The Dissertation Committee for Cédric Hervé David Certifies that this is the approved version of the following dissertation:

Towards river flow computation at the continental scale

Committee:

David R. Maidment, Supervisor

Zong-Liang Yang, Co-Supervisor

Randall J. Charbeneau

Omar Ghattas

Ben R. Hodges

Loukas F. Kallivokas

Towards river flow computation at the continental scale

by

Cédric Hervé David, M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2009

Dedication

To my family, for their love and support.

Acknowledgements

This work was partially supported by the U.S. National Aeronautics and Space Administration under the Interdisciplinary Science Project NNX07AL79G, by the U.S. National Science Foundation under project EAR-0413265: CUAHSI Hydrologic Information Systems, by Ecole des Mines de Paris, France, by the American Geophysical Union under a Horton (Hydrology) Research Grant and by the University Corporation for Atmospheric Research under the Graduate Student Visitor Program.

I am deeply grateful to my PhD advisors, Dr. David Maidment and Dr. Zong-Liang Yang. I sincerely appreciate that they allowed me the freedom to work on a subject of interest to me while supporting, guiding and challenging my endeavors. I have learned a considerable amount from them and earnestly hope that this dissertation is worthy of the trust they granted me.

I would like to thank the other members of my PhD committee: Dr. Randall Charbeneau, Dr. Omar Ghattas, Dr. Ben Hodges and Dr. Loukas Kallivokas for their continued advice on this work.

At Ecole des Mines de Paris, I am very grateful to Dr. Florence Habets for allowing me to visit her team and for broadening my scientific horizons. I would also like to thank Mr. Pascal Viennot, Dr. Emmanuel Ledoux and Dr. Patrick Goblet.

I thank Dr. Joël Noilhan, Dr. Eric Martin and Dr. Jean-Francois Mahfouf from the French Centre National de Recherches Météorologiques, for their scientific encouragements and for our enlightening discussions.

At the U.S National Center for Atmospheric Research I am very grateful to Dr. David Gochis for hosting my visit, for introducing me to the concepts of land surface modeling and for fruitful collaborations. I would also like to thank Dr. Larry Winter, Dr. Olga Wilhelmi and Dr. Jennifer Boehnert.

I thank Dr. Jeff Weber, Dr. Ben Domenico and Mr. John Caron of Unidata for their guidance through the maze of meteorological datasets.

At the Texas Advanced Computing Center, I thank Dr. Victor Eijkhout and Dr. Karl Schulz for their help with parallel computing and the center itself for providing computer resources.

I would like to express my gratitude to Mrs. Sylviane Wignacourt, of the French Ecole Centrale de Lille, who supported my application to The University of Texas at Austin and without whose trust this experience would not have been possible.

Towards river flow computation at the continental scale

Publication No._____

Cédric Hervé David, Ph.D.

The University of Texas at Austin, 2009

Supervisor: David R. Maidment

Co-Supervisor: Zong-Liang Yang

The work presented in this dissertation informs on river network modeling at large scales using geographic information systems, parallel computing and the latest advancements of atmospheric and land surface modeling. This work is motivated by the availability of a vector-based Geographic Information System dataset that describes the networks of streams and rivers in the United States, and how they are connected. A land surface model called Noah-distributed is used to provide lateral inflow to an NHDPlus river network in the Guadalupe River Basin in Texas. Challenges related to the projection of gridded hydrographic data from a coordinate system to another are investigated. The different representations of the shape of the Earth used in atmospheric science (spherical) and hydrology (spheroidal) can lead to a significant North-South shift on the order of 20 km at mid latitudes. A river network model called RAPID is developed and applied in a four-year study of the Guadalupe and San Antonio River Basins in Texas using the river network of NHDPlus. Gage measurements are used to estimate flow wave celerities in a river network and to assess the quality of RAPID flow

computations. The performance of RAPID in a massively-parallel computing environment is tested and further investigation of its scalability is needed before using RAPID at the state or federal level. The replacement by RAPID of the river routing scheme used in SIM-France – a hydro-meteorological model – is investigated in a ten-year study of river flow in France. While the formulation of RAPID improves the functionality of SIM-France, the flow simulations are comparable in accuracy to those previously obtained by SIM-France. Sub-basin parameterization was found to improve model results. A single criterion for quantifying the quality of river flow simulations using several river gages globally in a river network is developed that normalizes the square error of modeled flow to allow equal treatment of all gaging stations regardless of the magnitude of flow. The use of this criterion as the cost function for parameter estimation in RAPID allows better results than by increasing the degree of spatial variability in model parameters.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. BACKGROUND AND MOTIVATION

Hydrology is concerned with describing the motion of the waters of the Earth through the hydrologic cycle. This involves both the vertical exchange of water between the land surface with the atmosphere, and the horizontal movement of water through the landscape in surface and groundwater systems. Over the past two decades, the vertical water exchange between the land surface and the atmosphere has been extensively studied at the regional and the continental scale, but the horizontal movement of water in surface and groundwater systems at the regional and continental scale has not received an equivalent amount of attention.

The vertical interaction of the land surface and atmosphere is well described by gridded cell models of the landscape, but large scale studies of stream and river flow, such as the national FEMA Flood Map Modernization effort (http://www.fema.gov/plan/prevent/fhm/mm_main.shtm), are focused on applying the equations of one-dimensional flow to mapped river and stream reaches expressed as vector line objects. There is thus a mismatch between the forms of the spatial representations of vertical and horizontal water movement.

Hydrologic science is less advanced than atmospheric science and oceanography when considering its ability to understand, model and predict the water cycle on a large spatial scale and a long time scale. In the United States, the thirteen National Weather Service River Forecast Centers are each responsible for a different set of basins, and each modeler calibrates various models to these basins, without consistency across the nation [*Welles, et al.*, 2007]. Hydrologic science needs to be able to model river flow at the continental-scale using calculation techniques comparable to that of other geosciences. A

river flow model for large scale would enhance geosciences in an interoperable way and contribute to the modeling of a dynamic Earth. There are challenges in interoperability of geosciences. For example, complications due to whether a spherical or spheroidal shape is used for representation of the Earth can be important when linking atmospheric and hydrologic modeling at the continental scale. Furthermore, the concept of a boundary differs among the geosciences. Watersheds are of little importance for a global circulation atmospheric model.

Today, weather models, atmospheric reanalyses and observation datasets are available for the United States. The recent release of the enhanced version of the national hydrography dataset (NHDPlus) provides the hydrologic science community with unprecedented information about the nation's stream network (including connectivity information, slope, mean flow values, etc). Furthermore, supercomputing facilities are becoming increasingly powerful and accessible in the United States as NSF invests more in petascale computing facilities. The latest technological advancements in the United States offer new possibilities for development of next-generation river flow models.

The development of a river routing model should benefit from previous scientific work in other countries. In particular, SIM-France [*Habets, et al.*, 2008], the operational river forecasting system at the French national weather service has benefited from more than ten years of experience and provides insights for a next-generation river model.

Developing a framework for river flow calculation for the continental U.S. may lead, with supercomputing power increasing, to river flow computation for all rivers of the continental U.S. Such a tool will be beneficial to many types of studies such stream pollution, ecology of streams, coastal studies requiring river inflows, evaporation – precipitation – runoff budgets, etc.

1.2. OBJECTIVES

The present research focuses on improving methods for large-scale river routing by developing a river flow model for a network of thousands of connected stream reaches to be used within a framework consisting of one-way atmosphere/land surface model/river model coupling.

The objective of this dissertation is to address the following three sets of research questions:

1) How can the Noah-distributed land surface model be used to provide lateral and sub-surface water inflow to an NHDPlus river network? What are the potential sources of error in this process? What are the challenges related to the interoperability of land surface models and river networks?

2) How can a parallel computing river network model be developed to simultaneously compute the flow and volume of water in all river reaches of a vector stream network? How can model parameters be estimated automatically for an entire river network? How can this river model be applied to the NHDPlus dataset?

3) How can the river model be adapted to the quad-tree gridded river network of the SIM-France modeling framework and how does it compare with the existing river routing model? How can a criterion be developed to globally quantify the quality of river flow computations in a river network using a network of flow gages where the magnitude of the flow differs greatly among the gages?

1.3. OUTLINE OF DISSERTATION

This dissertation consists of a series of three related papers presented respectively in Chapter 2, Chapter 3 and Chapter 4. In the first paper, issues related to the coupling of land surface models with river networks are examined. In particular, errors due to the shape of the Earth and resulting interpretations of latitude are quantified; and technical

challenges of projecting hydrologically-based gridded datasets from a spherical Earth to a spheroidal are investigated. In the second paper, a river network routing model called RAPID is presented, which uses parallel computing and offers an automatic parameter estimation procedure. RAPID employs runoff calculations from the Noah land surface model with multi-physics options and is applied to the NHDPlus river network of the Guadalupe and San Antonio River Basins in Texas. In the third paper, a component of MODCOU, the current river routing model used in SIM-France, is replaced by RAPID. This replacement is motivated by the ability to directly calculate the flow and volume of water for all reaches of a river network and by the flexibility that this type of computation offers. A ten-year study over France is presented where river routing with RAPID and MODCOU are compared.

Chapter 2: Connecting a distributed land surface model with a vector-based river network[1]

2.1. ABSTRACT

The National Elevation, Hydrography and Land Cover datasets of the United States have been synthesized into a geospatial dataset called NHDPlus which is referenced to a spheroidal Earth, provides geospatial data layers for topography on 30-m rasters, and has vector coverages for catchments and river reaches. In this paper, we examine the integration of NHDPlus with the Noah-distributed model. In order to retain compatibility with atmospheric models, Noah-distributed utilizes surface domain fields referenced to a spherical rather than spheroidal Earth in its computation of vertical land surface/atmosphere water and energy budgets (at coarse resolution) as well as horizontal cell-to-cell water routing across the land surface and through the shallow subsurface (at fine resolution). Two data-centric issues affecting the linkage between Noah-distributed and NHDPlus are examined: 1) the shape of the Earth and 2) the linking of gridded landscape with a vector representation of the stream and river network. At mid-latitudes the errors due to projections between spherical and spheroidal representations of the Earth are significant. A catchment-based "pour point" technique is developed to link the raster and vector data to provide lateral inflow from the landscape to a one-dimensional river model. We conclude that, when Noah-distributed is run uncoupled to an atmospheric model, it is advantageous to implement Noah-distributed at the native spatial scale of the digital elevation data and the spheroidal Earth of the NHDPlus dataset rather

[1] Substantial portions of this chapter were first published in David, C. H., D. J. Gochis, D. R. Maidment, W. Yu, D. N. Yates, and Z.-L. Yang (2009), Using NHDPlus as the Land Base for the Noah-distributed model, *Transactions in GIS, (accepted for publication)*.

than transforming the NHDPlus dataset to fit the coarser resolution and spherical Earth shape of the Noah-distributed model.

2.2. INTRODUCTION

Hydrology is concerned with both the vertical exchanges of water between the land surface and the atmosphere, and the horizontal movement of water over and through the landscape in surface and groundwater systems. Over the past several decades, both vertical and horizontal water movement have been studied extensively though, generally, in relative isolation from one another [c.f. *Lyon, et al.*, 2008]. Specifically, the land surface models (LSMs) used as lower boundary conditions for numerical weather prediction models and global climate models focus on the vertical exchanges of entities including water [*Yang*, 2004]. In LSMs, horizontal exchanges of water at the grid or subgrid scales are not usually considered. The vertical interaction of the land surface and atmosphere is reasonably well described by gridded models of the landscape. However, many basin scale studies of stream and river flow, such as the national FEMA Flood Map Modernization effort (http://www.fema.gov/plan/prevent/fhm/mm_main.shtm), apply the equations of one-dimensional open channel hydraulics to mapped river and stream reaches treated as distinct linear entities or vector objects. Additionally, highly irregular objects such as stream channels, groundwater basins, watersheds, lakes and reservoirs may be significantly mis-represented or wholly absent within comparatively coarsely-resolved gridded modeling domains whose grid sizes lack sufficient horizontal resolution to properly define the boundaries of important hydrographic features. The main objective of this paper is to present a modeling framework using a standardized LSM for numerical weather at the continental-scale with a high-resolution hydrographic database as its land information base. Issues related to the coupling of regular gridded modeling domains used in LSMs with spatially irregular river systems represented as vector objects are

investigated on the Guadalupe River Basin in Texas as a case-study example (see Figure 1).

2.2.1. Geometry of surface routing in hydrologic modeling

In this study, we focus on those types of hydrologic models which emphasize coupling to a river hydraulics, river dynamics, or, more commonly, a river routing module. Many existing hydrologic models do not use vector-based mapped rivers for their hydraulic modeling, opting for a gridded representation of the river network instead. Such models include CASC2D [*Julien, et al.*, 1995], LISFLOOD [*De Roo, et al.*, 2000], MODCOU [*Ledoux, et al.*, 1989], NWSRFS [*NOAA*, 2005] and MIKE-SHE [see, for example: *Sahoo, et al.*, 2006]. Another class of 'watershed' models, such as, TOPMODEL [*Beven and Kirkby*, 1979], Maidment et al. [1996] or Olivera and coworkers [*Olivera and Maidment*, 1999; *Olivera, et al.*, 2000] typically route water across and through the landscape to a river basin outlet and do not explicitly model channel flow processes. Although gridded representations of the land surface are most common, they are not the only representation employed. Triangular cells, such as those within triangulated irregular networks (TINs), are also sometimes used to characterize the land surface [e.g. the tRIBS model, *Ivanov, et al.*, 2004]; triangular edges may be used to represent river links as in the coupled river network and groundwater flow model developed by Gunduz and Aral [2005].

In all of the models listed above, hydrographic features have been adapted to match model frameworks. However, recent development of high resolution vector-based hydrographic databases on a continental-scale such as NHDPlus [*USEPA and USGS*, 2007] or on a global-scale such as HydroSHEDS [*Lehner, et al.*, 2006] offer the potential for improved accuracy in the mapping of hydrographic features across continental domains. The focus of the present study is on issues related to using these vector-based

hydrographic datasets within hydrologic models. As shown subsequently, we suggest that retaining the spatial and geographic accuracy of underlying hydrographic data is advantageous to reducing potentially large georeferencing errors resulting from coupling hydrographic and meteorological data representations.

2.2.2. Coordinate systems for hydrologic modeling

Hydrologic models require both land-base information (terrain elevation, hydrographic features and networks) and atmospheric forcing (precipitation, specific humidity, temperature, air pressure, wind speed, downward radiation, etc.). Provided with these data, hydrologic models calculate evaporation, surface radiation exchanges, soil moisture, lateral overland and subsurface flow and streamflow. Commonly – and for mathematical simplicity – a Cartesian (x,y,z) coordinate system is used over the region of interest. Therefore, horizontal projections of geospatial data from a geographic coordinate system in latitude, longitude (λ, φ) to a projected coordinated system (x,y) are needed in order to transform the curved surface of the Earth into the Cartesian coordinate system. A projected coordinate system consists of an Earth datum and a projection to transform the globe to a Cartesian coordinate system on a flat map; these transformations are typically performed within a geographic information system (GIS).

Hydrologic modeling has mainly been done either at the small watershed-scale or at the larger continental-scale, and the geographic coordinate systems for data at these two scales are different. As high-resolution datasets become increasingly available, the geo-referencing of data becomes important in ways that have not been as apparent earlier.

In watershed scale models, land surface data usually come from local high-resolution databases and atmospheric data are interpolated from local point measurements, which are more resolved (spatially and temporally) than results from atmospheric circulation models at these scales. In this case, the high-resolution land-base

is readily available in projected coordinates of a spheroidal Earth datum and meteorological stations are located using geographic coordinates on the same spheroidal Earth datum.

Another class of regional or continental hydrologic models is sometimes referred to as 'macro' scale hydrologic models following Shuttleworth [1988]. These include ISBA [*Noilhan and Planton*, 1989], Noah [*Ek, et al.*, 2003], and the Community Land Model [*Dickinson, et al.*, 2006] and have been developed by the atmospheric science community for coupling with atmospheric circulation models. The VIC model [*Liang, et al.*, 1994; *Liang, et al.*, 1996] also falls under this category although it is typically run in an uncoupled mode for hydrologic simulation and prediction. The purpose of these models is to provide an atmospheric circulation model with lower boundary conditions for energy and water balances at the surface. However, at regional or continental scales, streamflow calculation has rarely been a principle objective for macroscale hydrologic models (with the exception of VIC), but streamflow data are used to verify or improve the volumetric water balance of land surface and atmospheric circulation models. Because these regional/continental-scale models are coupled to atmospheric models, their grids are typically cast in a coordinate system which utilizes a spherical Earth datum (see Section 3.1.), as opposed to the more appropriate spheroidal datum.

Datum issues and associated spatial errors in geographic transformations have not been focused upon in coupling land surface/atmospheric circulation models, perhaps because the grid resolution of atmospheric models has been much coarser than the positional errors associated with geographic transformations. However, atmospheric models and spatially continuous data products are now available for the continental U.S at very high spatial resolutions. For instance, results from real time runs of the Weather Research and Forecasting model are available on a 12-km grid

(http://motherlode.ucar.edu:8080/thredds/catalog.html) and remotely sensed NEXRAD precipitation is available on a 4-km grid (Stage IV NEXRAD data are available at http://nomads.ncdc.noaa.gov/thredds/catalog.html). The increased geospatial resolution and quality of available atmospheric datasets suggests that datum and projection issues should not be neglected.

Therefore datum issues become important as high resolution atmosphere and land surface modeling is merged with high resolution hydrographic datasets, as is demonstrated in this study.

2.3. FRAMEWORK DESCRIPTION

A hydrologically-enhanced form of the Noah LSM [Noah-distributed, *Gochis and Chen,* 2003] has been developed that allows for cell-cell routing of flow across and through the landscape. In this study, the NHDPlus dataset is used as the land-base for Noah-distributed. NHDPlus is a GIS dataset that links the National Hydrography Dataset description of the mapped streams and water bodies of the nation with small catchments delineated around each stream reach. The Guadalupe Basin in Texas has about 3000 river and stream reaches and their surrounding catchments in the NHDPlus dataset (Figure 2). This basin was chosen for study because it has significant contributions to surface water flow from groundwater sources, because it has a large reservoir, Canyon Lake (surface area of 33 km^2), where the effect of reservoir releases on downstream flow dynamics has to be considered, and because it flows out into an important estuarine system at San Antonio Bay. Figure 3 shows three components of the geospatial framework used in this study. A schematic of processes in Noah LSM [*Chen, et al.*, 1996] is presented in Figure 3a), the overland and subsurface routing functionalities of Noah-distributed [*Gochis and Chen*, 2003] are in Figure 3b), and the river network as defined in NHDPlus for the Guadalupe River Basin in Texas is in Figure 3c).

2.3.1. NHDPlus as the land base for Noah-distributed

NHDPlus [*USEPA and USGS*, 2007] is a hydrologically enhanced land database that incorporates many of the best features of the National Hydrography Dataset (NHD), the National Elevation Dataset (NED), the Watershed Boundary Dataset (WBD) and the national Land Cover Dataset (NLCD). NHDPlus includes a stream network based on the medium resolution NHD (1:100,000 scale), explicit stream networking, feature characterization, and a number of additional attributes such as divergence, network

connectivity, stream order, and mean annual flow. Therefore, NHDPlus is a geospatial dataset that connects the land and water systems of the United States. A higher resolution NHD stream network at 1:24,000 scale exists for the United States but is not connected to the landscape by reach catchments as is NHDPlus. Within NHDPlus, the continental U.S. is divided in 18 regions, with their corresponding two-digit Hydrologic Unit Code (HUC, from 01 to 18). Data for Alaska is not available. Data for Hawaii is available within region 20. The Texas Gulf, within which the Guadalupe River Basin resides, is region 12 in NHDPlus.

In NHDPlus, the Digital Elevation Model (DEM) has been modified from the national elevation dataset to conform to the river network and the watershed boundary. Using the AGREE-DEM method [*Hellweger and Maidment*, 1997] for the river network, DEM walls are created at known watershed boundaries from the watershed boundary dataset similarly to Moore et al. [2004]. As a result, river and watershed delineation based on the modified DEM and its associated flow accumulation and direction grids conforms to available vector stream and watershed data. The spatial resolution of the raster datasets is 30 m in the NHDPlus DEM, flow direction and flow accumulation grids. NHDPlus rasters use the USGS national Albers projection with a spheroidal Earth, defined by the North American Datum of 1983 [*Schwarz and Wade*, 1990].

NHDPlus uses connected river reaches, and for each reach a catchment is defined to delineate its local drainage area. This reach catchment is assigned a unique identifier, the COMID, and all features and attributes pertaining to this reach and its catchment are labeled similarly. Within NHDPlus, the river network value-added attribute table includes FromNode and ToNode fields that can be used to specify how streams and reaches are connected to form the river network. The NHDPlus dataset also has an estimated slope and mean annual flow for each river reach. For the continental United

States, the NHDPlus dataset has about 3 million stream and river reaches of average length 2 km, and the average catchment size defined around them is 3 km^2 in area. The Guadalupe River Basin has about 3,000 stream and river reaches of average length 3 km and the average catchment size is 5 km^2 in area.

NHDPlus also has integrated the National Land Cover Dataset (NLCD) and has as attributes the calculated percentage of coverage of each classified NLCD land cover (water, developed, barren, forested shrub land, etc.) for each catchment. While important, land cover is not the focus of the present study.

2.3.2. The Noah-distributed Model

Within our framework, the core physical model governing the one-dimensional (1D) vertical fluxes of energy and moisture is the Noah LSM. The one-dimensional Noah model simulates liquid and frozen soil moisture, soil temperature, skin temperature, snowpack depth and water equivalent, canopy water content, and energy and water fluxes at Earth's surface [*Mitchell,* 2005]. The Noah model has a long history, with successive versions extensively tested and validated, most notably within the Project for Intercomparison of Land surface Parameterizations [*Henderson-Sellers, et al.*, 1993], the Global Soil Wetness Project [*Dirmeyer, et al.*, 1999], and the Distributed Model Intercomparison Project [*Smith, et al.*, 2004]. Existing gridded versions of the 1D Noah model are coupled to real-time weather forecasting models such as the NCAR/Penn State University (NCAR/PSU) MM5, the advanced Weather Research and Forecasting (WRF) numerical weather prediction model, and the NCEP North American Model (an alternate version of WRF), which is used for performing operational weather prediction for the United States.

In Noah-distributed, a flow-routing-capable version of Noah [*Gochis and Chen,* 2003], the overland flow routing is calculated as a fully unsteady, explicit, finite

difference, one- or two-dimensional diffusive wave flowing over the land surface, similar to that used in the CASC2D model of [*Julien, et al.*, 1995]. 'Shallow' groundwater flow (down to 2-m depth) is also explicitly modeled using a quasi-steady state saturated flow model adapted from Wigmosta et al. [1994]. The horizontal flow into a stream network calculated by Noah-distributed is then a combination of surface runoff and shallow groundwater flow. More recently a baseflow module was also implemented within Noah-distributed which employs a simple bucket model to estimate time-evolving baseflow in perennial streams. Flow from the stream back to the landscape or aquifer is currently neglected. Although Noah-distributed also has the ability of routing water within river networks and through reservoirs, these calculations are not used in the present study.

The spatial resolution of the 1D Noah model is currently limited by the spatial resolution of land surface characterization (e.g. soils and vegetation) datasets. Therefore a subgrid modeling approach is used in which the vertical fluxes and land-atmosphere exchanges within Noah are calculated using gridcells on the order of 1 km x 1 km while a much finer grid, on the order of 100 m, is typically used for routing runoff over and through complex landscapes. This subgrid routing functionality is intended to build upon highly-resolved terrain datasets, such as the NHDPlus, and the need for adequately resolving terrain slopes.

2.4. LINKING A LAND SURFACE MODEL WITH A VECTOR-BASED RIVER NETWORK

2.4.1. Shape of the Earth

Most large scale atmospheric datasets that are available for North America are referenced on a Cartesian coordinate system projected from a spherical Earth. Such is the case for the Rapid Update Cycle and the North American Mesoscale model outputs, as well as for the North American Regional Reanalysis and the national aggregation of NEXRAD data. These data are all available online at http://nomads.ncdc.noaa.gov/thredds/catalog.html and http://motherlode.ucar.edu:8080/thredds/catalog.html. Most hydrologic datasets use a more accurate spheroidal Earth geometry. Vector data in NHDPlus are presented in geographic coordinates using the spheroidal NAD83 datum. This datum and the USGS national Albers projection is used for the NHDPlus rasters (DEM, flow direction and flow accumulation). The spheroidal or ellipsoidal shape used in NAD83 is that of the Geodetic Reference System 1980 [GRS80, *Moritz*, 1980]. To distinguish between a spherical and spheroidal Earth, two types of latitudes are needed: geocentric and geodetic as shown in Figure 4. Longitudes are not affected by this difference in Earth shape because it only involves North-South flattening. The geocentric latitude Φ is the acute angle measured between the equatorial plane and a line joining the center of the Earth and a point on the surface of the sphere or spheroid. The geodetic latitude φ' is the acute angle between the equatorial plane and a line drawn perpendicular to the tangent plane of a point on the reference sphere or spheroid. Normal map coordinates are given in longitude and geodetic latitude. On a sphere, geocentric and geodetic latitudes are equal. For the GRS80 spheroid, the semimajor axis is a=6378137m, the semiminor axis is b=6356752.3141m and the mean radius is $R1 = \frac{2a+b}{3} = 6371008.7714\text{m}$ [*Moritz*, 1980].

Gates [2004] derived the equations of atmospheric motion and in spheroidal coordinates and gave the following expression for the difference between geocentric and geodetic latitudes:

$$\delta = \varphi' - \Phi = \tan^{-1}\left(\frac{\sin(\varphi)\cdot\cos(\varphi)}{\sinh(\xi)\cdot\cosh(\xi)}\right) \tag{1}$$

where:

$$\varphi = \tan^{-1}\left(\tan(\varphi')\cdot\tanh(\xi)\right) \tag{2}$$

and:

$$\tanh(\xi) = \frac{b}{a} \tag{3}$$

where φ is the angle of the cone that is asymptotic to the hyperboloid orthogonal to the spheroid, and ξ is a dimensionless parameter.

Equations (1) to (3) applied to the GRS80 spheroid at mid latitude ($\varphi' = 45°$) give $\delta = 0°11'33"$, which corresponds to 21.4 km on the surface of a sphere (with the radius being the mean radius of the GRS80 spheroid). Therefore, at mid-latitudes, the degree of error that results from ignoring the different shapes of the Earth is on the order of 20 km. This error is 18.5 km at $\varphi' = 30°$, 18.6 km at $\varphi' = 60°$ and goes to zero at the equator and at the poles. Figure 5 shows the distance between points having the same numerical value for latitude depending on whether it is geocentric-based or geodetic-based, as a function of geodetic latitude. For each geodetic latitude, the corresponding geocentric latitude was calculated using the GRS80 spheroid and equations (1) to (3). The angular

difference was then multiplied by the mean radius of the GRS80 spheroid to determine a distance in kilometers. Comparable values were given by Van Sickle [2004].

Using the geocentric latitude of the sphere as the geocentric latitude on the spheroid is equivalent to performing a geocentric projection. Figure 6 shows the principle of the geocentric conversion, where the geocentric latitude on a sphere is taken as the geocentric latitude on a spheroid allowing the projection of a point M_{sphere} at the surface of the sphere to the corresponding location $M_{spheroid}$ on the surface of the spheroid. Two domains resulting from two interpretations of latitudes are shown in Figure 7. These two domains were created using the same numerical values for longitude and latitude, but assigning the latitudes either to the geodetic latitudes ($\varphi' = 28.3°$ and $\varphi' = 30.4°$ for domain a) or to the geocentric latitudes ($\Phi = 28.3°$ and $\Phi = 30.4°$ for domain b). The shift is on the order of 20 km in the North-South direction. As illustrated in Figure 7, errors of these magnitudes can be particularly important for terrestrial hydrological applications such as flood prediction where positional errors of a few kilometers can produce pronounced differences on catchment scale runoff response.

2.4.2. Spatial discretization in Noah-distributed and the NHDPlus

In several current applications of the Noah-distributed model, the computational grid has been set up using 100-m DEM and flow direction grids and a 1-km atmospheric data forcing. This means that the vertical water balance computations are done on a 1-km grid and the horizontal flow routing is performed on a nested 100-m grid. The native resolution of NHDPlus rasters is 30 m, which provides better explicit representations of topographic features, terrain slopes in particular, and is therefore capable of better representing hydrologically important terrain gradients than a coarser 100-m grid.

When a 30-m DEM is re-sampled to a coarser resolution of 100 m, the correspondence of the DEM, and its flow accumulation and flow direction fields with the

original river reaches and catchment boundaries hydrologic features is lost, thereby requiring a re-definition of surface hydrographic features. The resulting catchments also change shape and, potentially, location, and the associated basin outlets or pour points also change location. The reprocessing of the raster data along with the re-specification of hydrographic features to coarser grids is very time consuming. The same arguments hold for projecting of the NHDPlus raster data from its spheroidal Earth coordinates to the spherical Earth coordinates used in Noah-distributed. Thus, from a hydrographic data perspective, there are distinct advantages to directly utilizing an integrated DEM-hydrography data set such as the NHDPlus which avoids reprocessing all of the geospatial data to a different spatial resolution. The principal disadvantage of using the native 30-m grid of the NHDPlus dataset is need for substantially greater computing resources and longer model run times for Noah-distributed, particularly for large simulation domains. Such highly resolved domains also create a burden for data storage, analysis and visualization.

Given these considerations we recommend that, from a spatial accuracy standpoint, it is advantageous to adopt the NHDPlus spatial framework and to adapt the Noah-distributed model to execute on the NHDPlus grid, rather than to resample the NHDPlus land surface dataset to a coarser resolution. Thus, for studies of the Guadalupe River Basin, Noah-distributed utilizes a 30-m resolution DEM for overland and subsurface routing. The corresponding resolution for the one-dimensional Noah model was chosen as 900 m instead of the previous 1 km to allow for integer conversion between the two grids. Therefore, each 900-m land-atmosphere cell is constituted of 30 x 30 surface routing grid cells. The methods by which soil moisture and ponded water are disaggregated from the Noah-distributed land model grid onto the routing subgrid are described in Gochis and Chen [2003].

2.4.3. Catchment Pour Points

A "pour point" technique is used to link NHDPlus catchments to the vector-based stream network. This requires determining the outlet location of the catchment corresponding to each river reach of a basin. This location is taken as the point with the highest flow accumulation value on the raster DEM grid within the catchment. This connection is facilitated by the presence in the NHDPlus raster dataset of a catchment raster whose zone identifier is the COMID value of the catchment. By searching within this zone for the cell of maximum flow accumulation, the pour point cell is identified at the catchment outlet, as shown in Figure 8 where the same catchment is shown in both the gridded and vector environments of NHDPlus. Following the flow direction grid, water is allowed to flow on and below the land surface of each catchment within the calculations of Noah-distributed and is accumulated at the pour point. This water is then specified as the inflow to the corresponding river reach. Figure 8 shows how vector data (river reaches) are connected to the pour point of the NHDPlus catchment, thus achieving a conceptual translation between vector-based and raster-based environments. Therefore, the pour point method allows the use of the gridded landscape of NHDPlus within the Noah-distributed model to simulate the horizontal movements of water, while remaining compatible with the NHDPlus streams and reaches that can then be used for routing within a vector-based river network. Hence, this study presents a way to provide lateral inflow of water from the land surface to the river network.

2.5. CONCLUSIONS

In this study, the advantages and disadvantages of alternatives for spatially connecting atmospheric model grids with those from catchment and river models using a standardized national GIS vector river and raster terrain dataset (NHDPlus) and a standard land surface/atmospheric model (Noah) are discussed. The different shapes of the Earth that are used in atmospheric science (spherical) and hydrology (spheroidal) can lead to two different interpretations of latitudes: geocentric or geodetic. A shift in the North-South direction on the order of 20 km at mid latitudes results from these two interpretations. The magnitude of this shift is comparable to the grid cell size of high-resolution atmospheric datasets available today. This discrepancy must be avoided by projections from one datum to another. It is advantageous to keep the original spatial resolution and datum of the NHDPlus, and to project and resample atmospheric data instead when using NHDPlus as the land base for the Noah-distributed model. In doing so, the original spatial resolution of the terrain rasters, the shape of hydrographic features, and the connectivity between catchments and river reaches from the NHDPlus dataset are preserved. The spatial resolution of the domain used for computation of the movement of water through the landscape to the river reaches is higher than that used in previous Noah-distributed studies, hence requiring a more intense computational demand. However, this demand can be met through recent advances towards petascale computing architectures now underway at major modeling centers around the world.

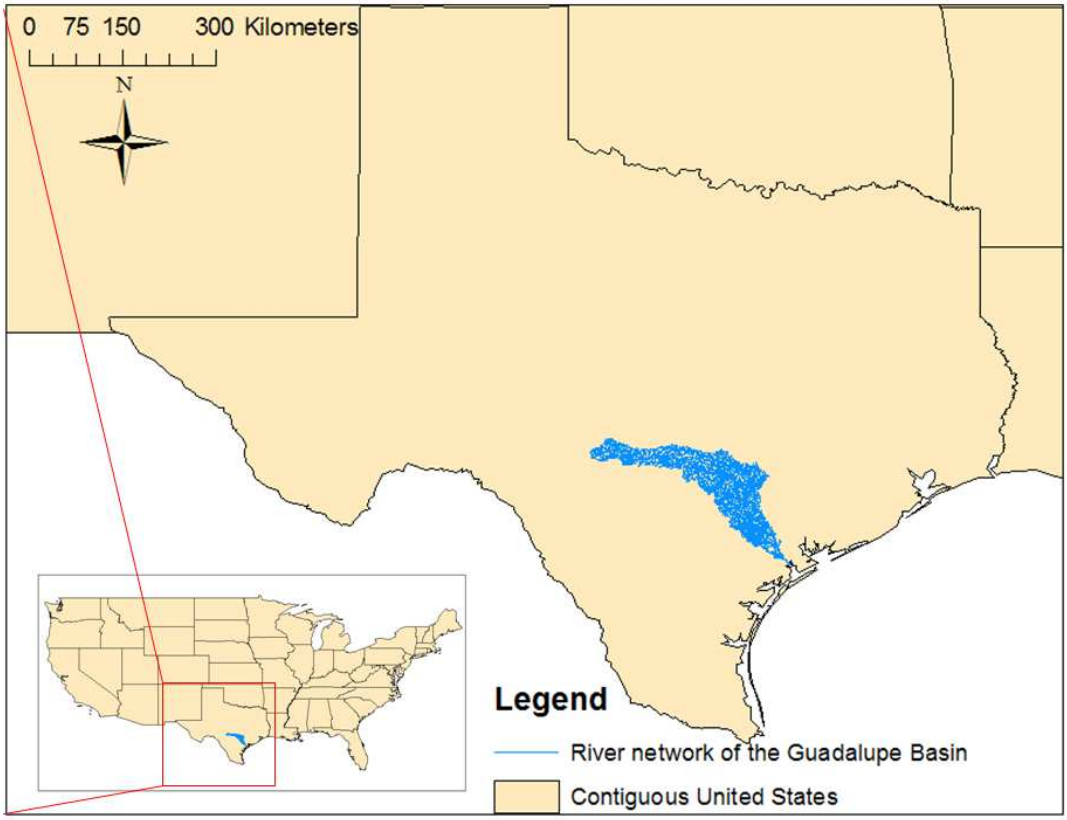

Figure 1 The Guadalupe Basin is located at the South-East of Texas

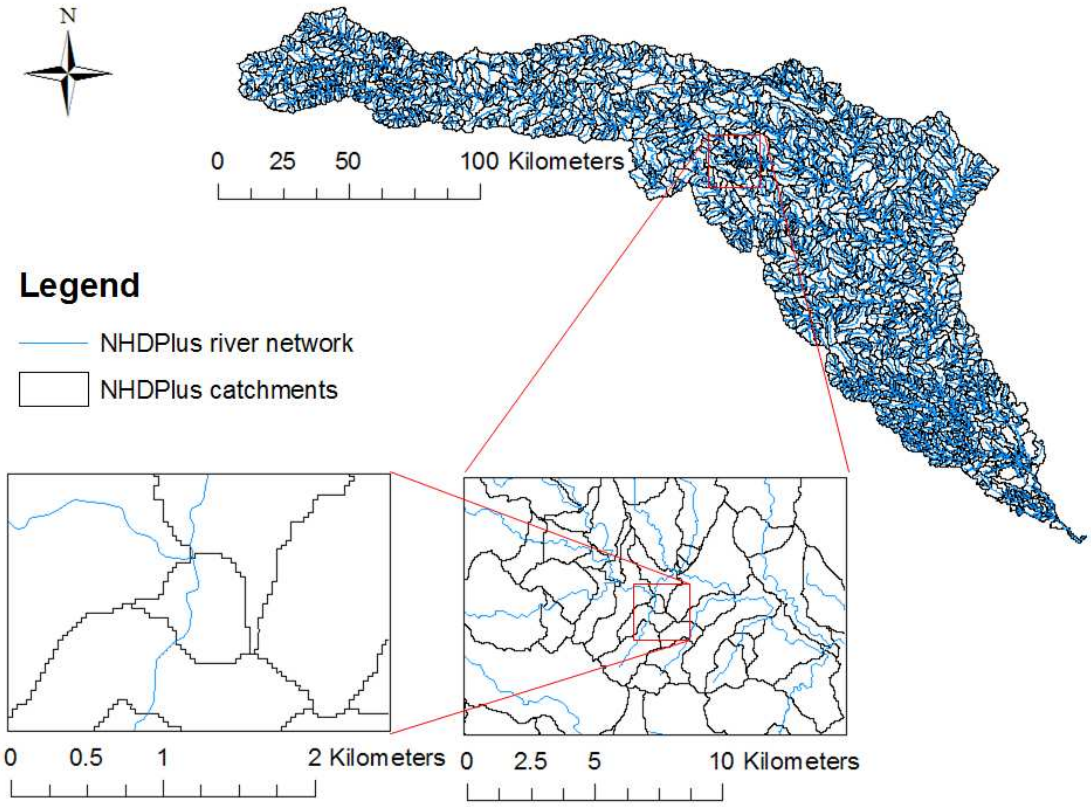

Figure 2 River and stream reaches and their surrounding catchments as defined in NHDPlus for the Guadalupe Basin

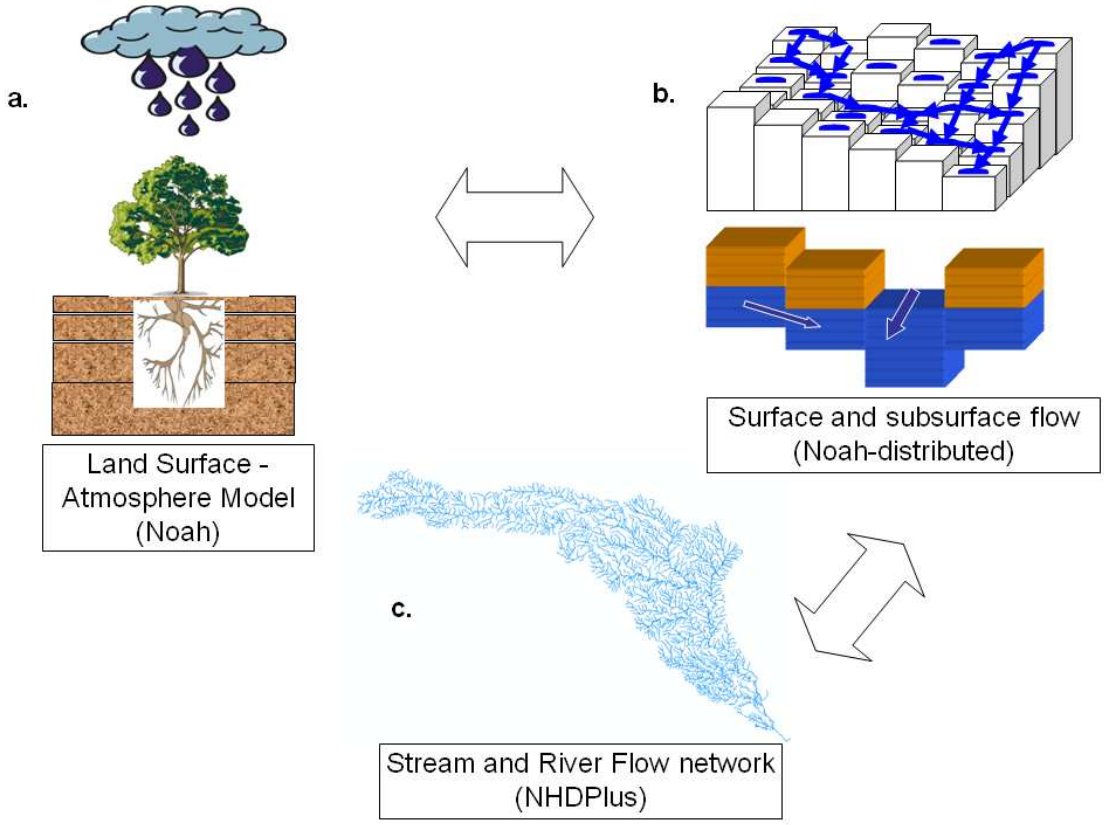

Figure 3 Components of the geospatial framework used in this study

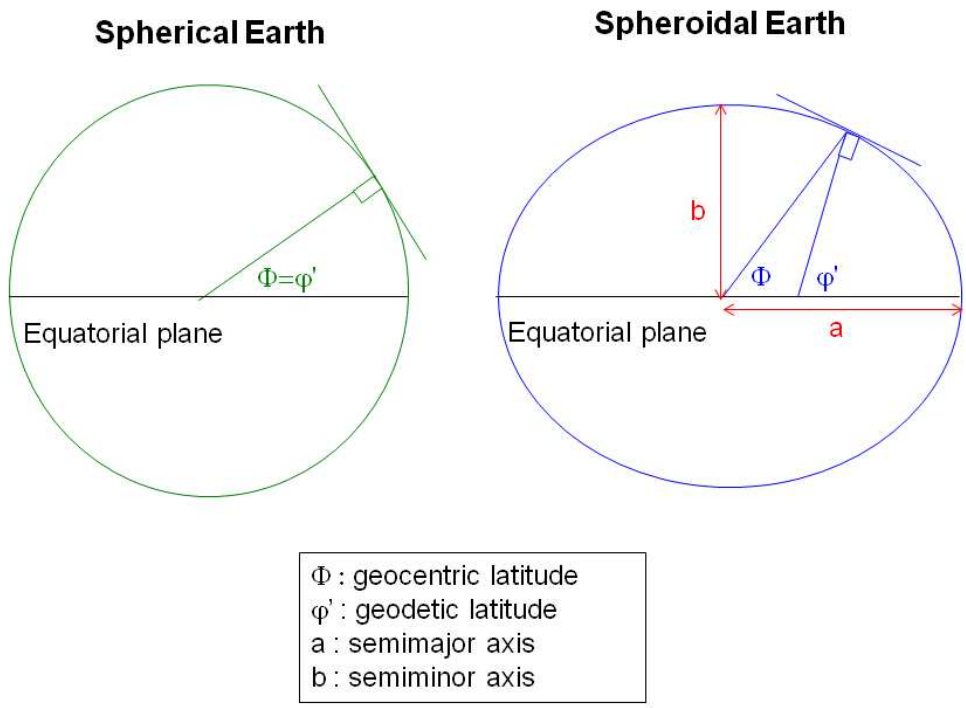

Figure 4 Geometry of spherical and spheroidal representations of the Earth.

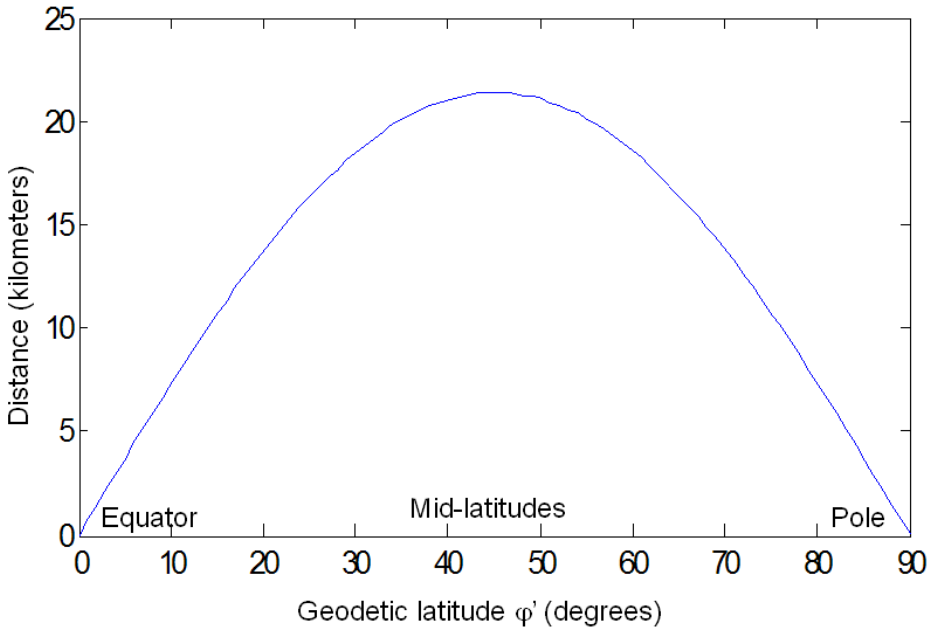

Figure 5 Distance between points having the same numerical value for latitude depending on whether it is geocentric-based or geodetic-based, as a function of geodetic latitude

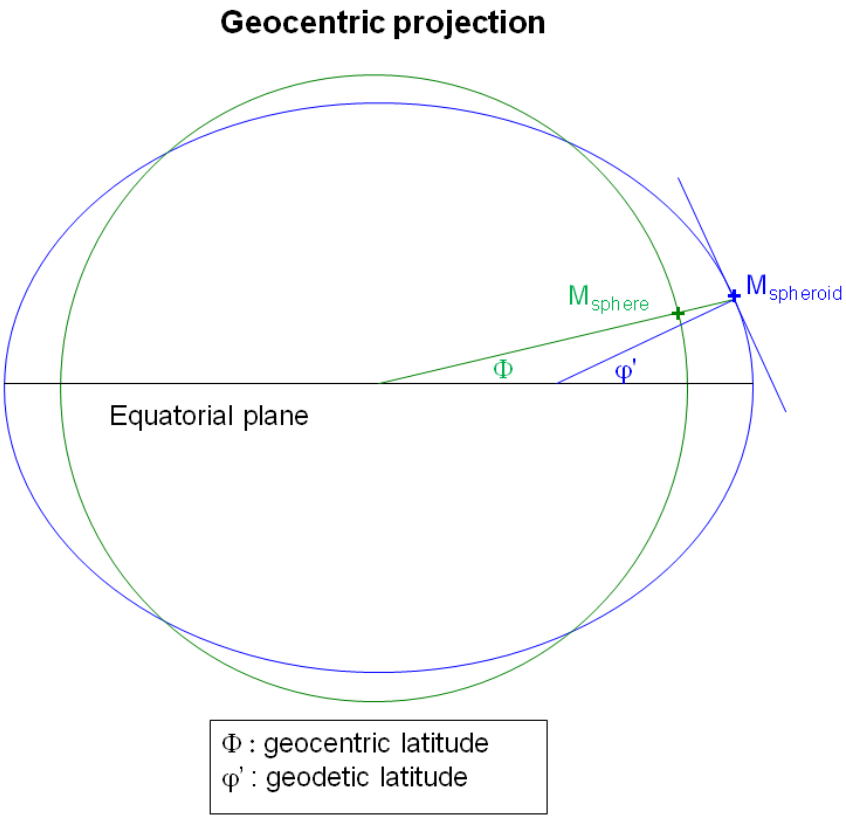

Figure 6 Geocentric projection of a sphere to a spheroid

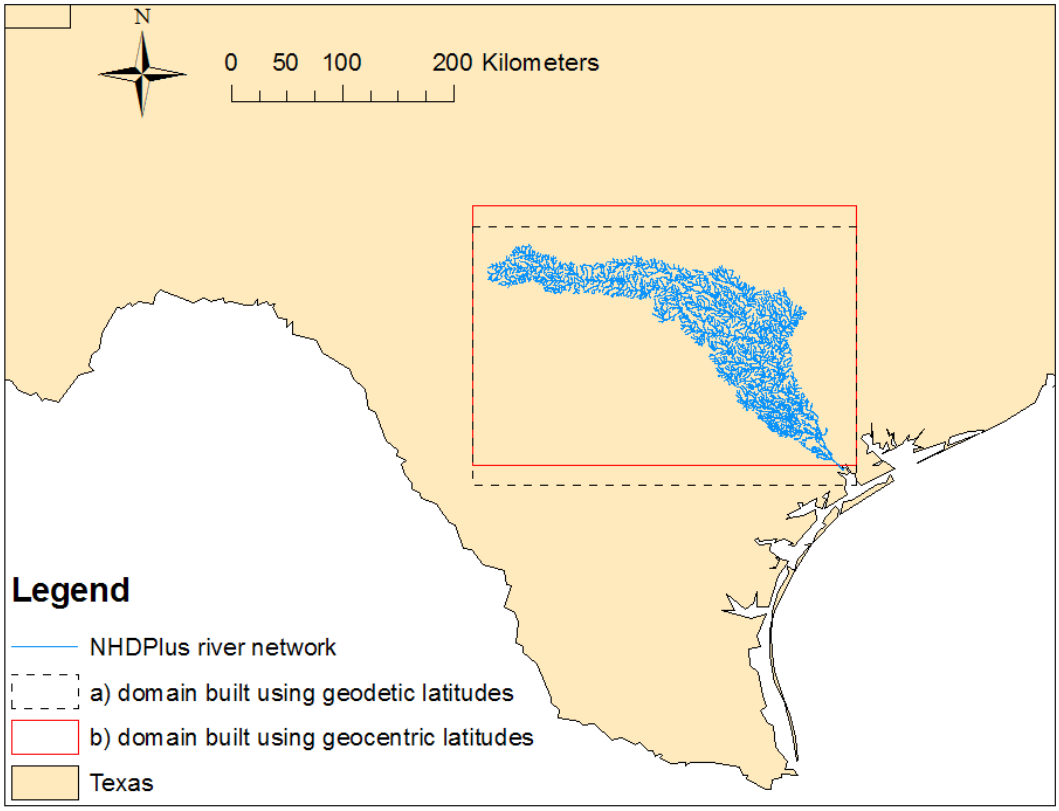

Figure 7 Different interpretations of latitude for the Guadalupe River Basin can lead to shifted locations of the same domain

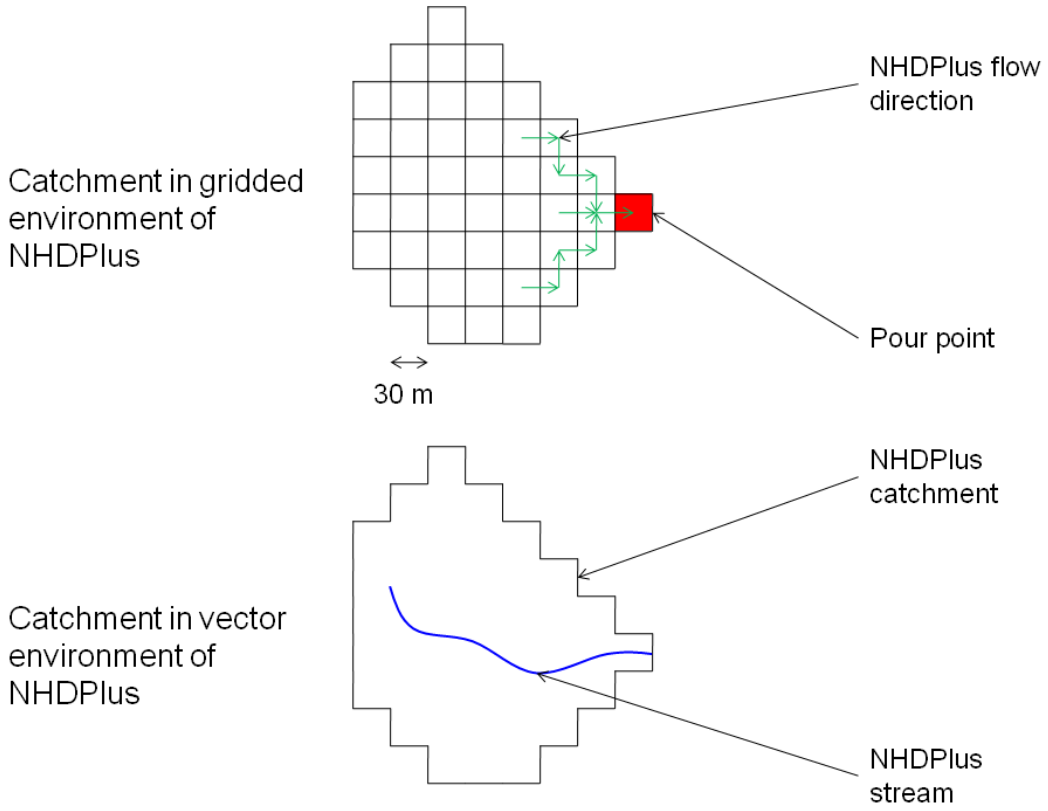

Figure 8 Connection between grid and vector environments of NHDPlus using pour points

Chapter 3: Routing application for parallel computation of discharge

3.1. ABSTRACT

The mapped rivers and streams of the contiguous United States are available in a geographic information system (GIS) dataset called NHDPlus. This hydrographic dataset has about 3 million river and water body reaches along with information on how they are connected into networks. A river network model called RAPID is developed for the NHDPlus river network and applied to the Guadalupe and San Antonio River Basins in Texas, whose lateral inflow to the river network is calculated by a land surface model. RAPID allows for a matrix-based calculation of flow and volume of water in all reaches of a river network, with many thousands of reaches. Gages from the USGS National Water Information System are used to assess the quality of the calculations with about 1 gage available for each 160 reaches simulated. Flow wave celerities are estimated at multiple locations of a basin using 15-minute data and can be reproduced reasonably with RAPID. RAPID can be adapted for parallel computing but challenges related to parallel computing are significant.

3.2. INTRODUCTION

Land surface models (LSMs) have been developed by the atmospheric science community to provide atmospheric models with bottom boundary conditions (water and energy balance) and to serve as the land base for hydrologic modeling in the context of general circulation models (GCMs). Over the past two decades, overland and subsurface runoff calculations done by LSMs have extensively been used to provide water inflow to river routing models that calculate river discharge [*De Roo, et al.*, 2003; 1999a; 1999b; 1999c; *Habets, et al.*, 2008; 1998b; 1998c; *Lohmann, et al.*, 2004; *Maurer, et al.*, 2001; *Oki, et al.*, 2001; *Olivera, et al.*, 2000]. However, river routing within LSMs has traditionally been done using gridded river networks that best fit the computational domain used in LSMs. Today, geographic information system (GIS) hydrographic datasets are increasingly becoming available at the continental scale such as NHDPlus [*USEPA and USGS*, 2007] and the global scale such as HydroSHEDS [*Lehner, et al.*, 2006]. These datasets provide a vector-based representation of the river network using the "blue line" mapped rivers and streams. The present study links a land surface model with a new river network model called RAPID using NHDPlus for the representation of the river network. The latest work on GCMs by the international scientific community, especially by the intergovernmental panel on climate change, opens potential studies of the evolution of water resources with global change. Using mapped streams and water bodies in LSMs could benefit the resulting assessment of the impact of global change in water resources.

All models and datasets used herein are available at least for the contiguous United States. The work presented here has focused on the Guadalupe and San Antonio Basins in Texas (see Figure 9) together covering a surface area of about 26,000 km^2.

These basins have about 5,000 river reaches and their corresponding catchments in the NHDPlus dataset (see Figure 10) out of 3 million for the United States. These basins are also chosen for study because of significant contributions to surface water flow from groundwater sources, because of a large reservoir, at Canyon Lake, where the impacts of constructed infrastructure on flow dynamics have to be considered, and because these rivers flow out into an estuarine system at San Antonio Bay.

The research presented in this paper aims at answering the following questions: how can a river model be developed for calculation of flow and volume of water in a river network of thousands of river reaches? How can the connectivity information in NHDPlus be used to run a river network model in part of the United States? How can flow at ungaged locations be reconstructed, and how can model computations be assessed based on available measurements? Can parallel computing be used for river flow modeling?

First, the model development is presented. Then, the modeling framework for the application river flow calculation in the Guadalupe and San Antonio River Basins is developed, followed by results and conclusions.

3.3. MODEL DEVELOPMENT

The model presented here is named RAPID (Routing Application for Parallel computatIon of Discharge). RAPID uses the Muskingum method that was first introduced by McCarthy [1938] and has been extensively studied in the literature in the past 70 years. Among the most notable papers related to the Muskingum method, Cunge [1969] showed the Muskingum method is a first order approximation of the kinematic and diffusive wave equation and proposed a variable-parameter Muskingum method, known as the Muskingum-Cunge method which itself is a second order approximation of the kinematic and diffusive wave equation. More recently, Todini [2007] developed a mass-conservative variable-parameter Muskingum method known as the Muskingum-Cunge-Todini method. As a first step, the original Muskingum method is used in this study because there are significant problems related to doing flow routing on NHDPlus networks that need be overcome before more sophisticated flow equations are used. However, the physics of flow could be improved with many variations based on the Muskingum method or adapted to include the Saint Venant equations.

3.3.1. Calculation of flow and volume of water in a river network

In a network of thousands of reaches, matrices are needed for network connectivity and flow computation. The backbone of RAPID is a vector-matrix version of the Muskingum method:

$$\left(\mathbf{I}-\mathbf{C}_1\cdot\mathbf{N}\right)\cdot\mathbf{Q}\left(t+\Delta t\right)=\mathbf{C}_1\cdot\mathbf{Q}^{\mathbf{e}}\left(t\right)+\mathbf{C}_2\cdot\left(\mathbf{N}\cdot\mathbf{Q}\left(t\right)+\mathbf{Q}^{\mathbf{e}}\left(t\right)\right)+\mathbf{C}_3\cdot\mathbf{Q}\left(t\right) \quad (4)$$

where t is time and Δt is the river routing time step. The bolded notation is used for vectors and matrices. $\mathbf{I}$ is the identity matrix. $\mathbf{N}$ is the river network matrix. $\mathbf{C_1}, \mathbf{C_2}$ and $\mathbf{C_3}$ are parameter matrices. $\mathbf{Q}$ is a vector of outflows from each reach, and $\mathbf{Q^e}$ is a vector of lateral inflows for each reach.

Equation (4) is used for river network routing and can be solved using a linear system solver. The vector-matrix notation provides one flow equation for the entire river network, therefore avoiding spatial iterations. For a river network with m river reaches, all vectors are of size m and all matrices are square of size m. Each element of a vector corresponds to one river reach in the network. For performance purposes, all matrices are stored as sparse matrices (only the non-zero values are recorded). A five-reach, two-node and two-gage river network is used here to clarify the mathematical formulation of the river network model and is shown in Figure 11a). The river network is made up of a combination of river reaches similar to that of Figure 11b). The model formulation is presented here for a small river network but can be generalized to any size of river network.

$\mathbf{Q}$ is a vector of the outflows Q_j of all reaches of the river network, where j is the index of a river reach within the network:

$$\mathbf{Q}(t) = \begin{bmatrix} Q_1(t) \\ Q_2(t) \\ Q_3(t) \\ Q_4(t) \\ Q_5(t) \end{bmatrix} = \left[Q_j(t) \right]_{j \in [1,m]} \tag{5}$$

$\mathbf{Q^e}$ is a vector of flows Q_j^e that are lateral inflows to the river network. Lateral inflows include runoff, groundwater or any type of forced inflow (outflow at a dam, pumping, etc.):

$$\mathbf{Q^e}(t) = \left[Q_j^e(t) \right]_{j \in [1,m]} \tag{6}$$

$\mathbf{Q^e}$ is provided by a land surface model, whose time step is coarser than the river routing time step. Two assumptions are made in the development of RAPID, one regarding the temporal variability of $\mathbf{Q^e}$ and one regarding the location at which $\mathbf{Q^e}$ enters the river network. In this study, the river routing time step is 15 minutes and inflow from land surface runoff is available every 3 hours. In the derivation of Equation (4), $\mathbf{Q^e}$ is assumed constant (i.e. $\mathbf{Q^e}(t+\Delta t) = \mathbf{Q^e}(t)$) over all 15-minute river routing time steps included within a given land surface model 3-hour time step. This temporal uniformity simplifies the river network model formulation, limits the quantity of input data and facilitates the coupling with land surface models. This assumption is valid at all times except at the last routing time step before a new $\mathbf{Q^e}$ is made available. Moreover, the external inflow $\mathbf{Q^e}$ is assumed to enter the network as an addition to the upstream flow. With these two assumptions, the Muskingum method applied to reach 5 in Figure 11b) gives the following:

$$\begin{aligned} Q_5(t+\Delta t) = {} & C_1 \cdot \left[Q_3(t+\Delta t) + Q_4(t+\Delta t) + Q^e{}_5(t) \right] \\ & + C_2 \cdot \left[Q_3(t) + Q_4(t) + Q^e{}_5(t) \right] \\ & + C_3 \cdot Q_5(t) \end{aligned} \tag{7}$$

where C_1, C_2 and C_3 are the well known Muskingum parameters that are stated in Equation (9). Equation (4) is a generalization of Equation (7) using a vector-matrix notation.

$\mathbf{N}$ is a network connectivity matrix. Berge [1958] proposed the concept of matrices associated with graphs. This concept can be applied to the river network in Figure 11a) in order to create the network matrix $\mathbf{N}$ given in Equation (8) in both full and sparse formats. The network connectivity matrix is a square matrix whose dimension is total number of reaches in the network. A value of one is used at row i and column j if reach j flows into reach i and zero is used everywhere else.

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} & & & & \\ & & & & \\ 1 & 1 & & & \\ & & & & \\ & & & 1 & 1 \end{bmatrix} \tag{8}$$

The upstream inflow to the network can therefore be computed by multiplying the network connectivity matrix $\mathbf{N}$ by the vector of outflows $\mathbf{Q}$. In case of a divergence in the river network (when going downstream) or in case of a loop, a unique reach (the major divergence) is used to carry all the upstream flow and the other reaches (minor divergences) carry only the flow that results from their lateral inflow. This formulation could be modified to take into account given fractions of flows that separate into different parts of a divergence if that information is available.

$\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$ are diagonal matrices with their diagonal elements being the coefficients used in the Muskingum method [*McCarthy*, 1938], respectively C_{1j}, C_{2j} and C_{3j} such that:

$$C_{1j} = \frac{\frac{\Delta t}{2} - k_j \cdot x_j}{k_j \cdot (1 - x_j) + \frac{\Delta t}{2}}, \quad C_{2j} = \frac{\frac{\Delta t}{2} + k_j \cdot x_j}{k_j \cdot (1 - x_j) + \frac{\Delta t}{2}}, \quad C_{3j} = \frac{k_j \cdot (1 - x_j) - \frac{\Delta t}{2}}{k_j \cdot (1 - x_j) + \frac{\Delta t}{2}} \tag{9}$$

where k_j is a storage constant (with dimension of a time) and x_j a dimensionless weighting factor characterizing the relative influence of the inflow and the outflow on the volume of the reach j. The Muskingum method is stable for any $x \in [0, 0.5]$, regardless of the value of k and Δt [*Cunge*, 1969]. For any j: $C_{1j} + C_{2j} + C_{3j} = 1$.

In RAPID, the parameters k and x of the Muskingum method are allowed to differ from one river reach to another, and corresponding vectors are defined in Equation (10):

$$\mathbf{k} = \left[k_j \right]_{j \in [1,m]} \quad , \quad \mathbf{x} = \left[x_j \right]_{j \in [1,m]} \tag{10}$$

The constants defined in Equation (9) are used as the diagonal elements of the matrices $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$. Equation (11) shows an example for $\mathbf{C_1}$. $\mathbf{C_2}$ and $\mathbf{C_3}$ are treated similarly.

$$\mathbf{C_1} = \begin{bmatrix} C_{1_1} & & & & \\ & C_{1_2} & & & \\ & & C_{1_3} & & \\ & & & C_{1_4} & \\ & & & & C_{1_5} \end{bmatrix} \tag{11}$$

The sum $\mathbf{C_1} + \mathbf{C_2} + \mathbf{C_3}$ equals the identity matrix.

The calculation of the volume of water in a given reach can be needed for coupling with groundwater models. Here, the first order, explicit, forward Euler method is applied to the continuity equation to calculate the volume of water in each river reach of the network, as shown in Equation (12) where the first, second and third terms of the right-hand-side are the volume of water that respectively were in the river reach, flowed into the reach, and discharged from the reach:

$$\mathbf{V}(t+\Delta t) = \mathbf{V}(t) + \left(\mathbf{N}\cdot\mathbf{Q}(t) + \mathbf{Q}^{\mathbf{e}}(t)\right)\cdot\Delta t - \mathbf{Q}(t)\cdot\Delta t \tag{12}$$

where $\mathbf{V}$ is a vector of the volume of water V_j in each river reach j:

$$\mathbf{V}(t) = \left[V_j(t)\right]_{j\in[1,m]} \tag{13}$$

3.3.2. Parameter estimation

In order to estimate the parameters $\mathbf{k}$ and $\mathbf{x}$ to be used in RAPID, an inverse method is developed. The principle of an inverse method is to optimize the parameters of

a model so that the outputs of the model approach observations. A cost function reflecting the difference between model calculations and observations is needed to assess the quality of a set of model parameters. The best set of parameters is chosen as the set that minimizes the cost function, and is determined through optimization. A square-error cost function ϕ is chosen:

$$\phi(\mathbf{k},\mathbf{x}) = \sum_{t=t_o}^{t=t_f} \left[\left(\frac{\overline{\mathbf{Q}}(t) - \mathbf{Q}^{\mathbf{g}}(t)}{f} \right)^T \cdot \mathbf{G} \cdot \left(\frac{\overline{\mathbf{Q}}(t) - \mathbf{Q}^{\mathbf{g}}(t)}{f} \right) \right] \tag{14}$$

where the summation is made daily. t_o and t_f are respectively the first day and last day used for the calculation of ϕ. The model parameter vectors $\mathbf{k}$ and $\mathbf{x}$ are kept constant within the temporal interval $[t_o, t_f]$, and the cost function is calculated several times with different sets of parameters during the optimization procedure. f is a scalar that allows ϕ to be on the order of magnitude of 10^1 which is helpful for automated optimization procedures. $\overline{\mathbf{Q}}(t)$ is the daily-average outflow vector, calculated based on the mean of all routing time steps in a given day. $\mathbf{Q}^{\mathbf{g}}(t)$ is a vector with the total number of river reaches for dimension, with the daily value observed $Q_j^g(t)$ corresponding to reach j where gage measurements are available, and zero where no gage is available. $\mathbf{G}$ is a sparse diagonal matrix that allows the dot-product to survive only where gages are available, so that $\mathbf{G}$ has a value of one on the diagonal element of index j if a gage is available on reach j and zero everywhere else. Using the example network given in Figure 11a), $\mathbf{G}$ and $\mathbf{Q}^{\mathbf{g}}(t)$ take the following form:

$$\mathbf{G} = \begin{bmatrix} & & \\ & 1 & \\ & & \\ & & 1 \end{bmatrix} \quad , \quad \mathbf{Q^g}(t) = \begin{bmatrix} 0 \\ 0 \\ \mathrm{Q}_3^{\mathrm{g}}(t) \\ 0 \\ \mathrm{Q}_5^{\mathrm{g}}(t) \end{bmatrix} \tag{15}$$

According to Fread [1993], $x \in [0.1; 0.3]$ in most streams. By analogy with the kinematic wave equation, Cunge [1969] showed that the parameter k of the Muskingum method is the travel time of a flow wave through a river reach. For a given river reach j of length L_j where a flow wave of celerity c_j travels, k_j is obtained by dividing the length by the celerity of the wave, as shown in Equation (16):

$$k_j = \frac{L_j}{c_j} \tag{16}$$

Although the routing model defined by Equation (4) allows for variability of the parameters (k_j, x_j) on a reach-to-reach basis, attempting to estimate model parameters independently for all the reaches of a basin would be a costly undertaking. Therefore, the search for optimal parameters is limited to determining two multiplying factors λ_k and λ_x such that:

$$k_j = \lambda_k \cdot \frac{L_j}{c_j} \quad , \quad x_j = \lambda_x \cdot 0.1 \tag{17}$$

At the end of the optimization procedure, one couple (λ_k, λ_x) is determined for a given basin in the network, as a first step.

3.3.3. Model implementation

The routing model is coded in Fortran 90 using the Portable, Extensible Toolkit for Scientific Computation (PETSc) mathematical library [*Balay, et al.*, 1997; *Balay, et al.*, 2001; *Balay, et al.*, 2004] and the Toolkit for Advanced Optimization (TAO) optimization library [*Benson, et al.*, 2007]. PETSc and TAO are built upon the Message Passing Interface [*Dongarra, et al.*, 1994] and allow for parallel computing. RAPID is run on single- and multiple-processor workstations as well as on Lonestar, a supercomputer running at the Texas Advanced Computing Center (http://www.tacc.utexas.edu/resources/hpcsystems/#lonestar). The linear system solver used in PETSc is a conjugate gradient squared solver, and the optimization method used in TAO is a line search algorithm called the Nelder-Mead method.

3.4. APPLICATION

RAPID is designed to handle large routing problems. Given a river network and connectivity information as well as lateral inflow to the river network, RAPID can run on any river network. In this study, a framework for computation of river flow in the Guadalupe and San Antonio River Basins is developed that uses a one-way modeling framework with an atmospheric dataset, a land surface model and RAPID as the river model. This section presents how NHDPlus connectivity can be utilized in RAPID, how a land surface model is used to provide lateral inflow to the river network, and the meteorological forcing prepared.

3.4.1 RAPID used on NHDPlus

NHDPlus [*USEPA and USGS*, 2007] is a geographic information system (GIS) database for the hydrography of the United States. This database provides the mapped streams and rivers as well as the catchments that surround them. NHDPlus is based on the medium resolution 1:100,000 scale national hydrographic dataset (NHD). One of the main improvements in NHDPlus is the network connectivity available in the value added attributes (VAA) table for the river network. Each NHDPlus reach in the national network is assigned a unique integer identifier called COMID. NHDPlus catchments also have a COMID, the same COMID being used for the reach and its local contributing catchment. Nodes are located at the two ends of each NHDPlus river reach. A unique integer identifier is given to all nodes in the national river reach network. The VAA table includes *FromNode* and *ToNode* fields that give which node is upstream and which is downstream of a given reach. Two reaches that are connected in a river network share a node, and the reach j flows into the reach i if $ToNode(j) = FromNode(i)$. The

NHDPlus connectivity between reaches, catchments and nodes is illustrated for three catchments of the Guadalupe and San Antonio basins in Figure 12.

In its current formulation, RAPID can handle several upstream reaches but only one unique downstream reach. However, divergences exist in mapped river networks, as they do in NHDPlus. The VAA table offers a *Divergence* field to each of the river reaches (with values of 0 – not part of a divergence, 1 – main path of a divergence, 2 – minor path of a divergence). In the current formulation of RAPID, the main part of a divergence carries all the upstream flow. The *FromNode*, *ToNode* and *Divergence* fields are used to populate the network matrix given in Equation (8), by means of the following logical statement:

$$\forall (i,j) \in [1,m]^2,\ \textit{if}\ \left[FromNode(i) = ToNode(j) \right]\ \textit{and}\ \left[Divergence(j) \neq 2 \right] \Rightarrow N_{i,j} = 1 \quad (18)$$

where $N_{i,j}$ is the element of $\mathbf{N}$ located at row i and column j. Therefore, upstream to downstream connection is conserved if the downstream reach is the major branch of a divergence or if it is not part of a divergence at all, but the connection is not made for a minor branch of a divergence.

The VAA table only has information for the river reaches whose flow direction is known. There are a total of 5175 river reaches with known direction within the Guadalupe and San Antonio river basins (as shown in Figure 10). These 5175 reaches have an average length of 3.00 km and the average catchment size defined around them is 5.11 km^2 in area and are all used for this study.

3.4.2. Land surface model and coupling with RAPID

Within this study, the core physical model governing the one-dimensional vertical fluxes of energy and moisture is the Community Noah Land Surface Model with Multi-Physics Options, hereafter referred to as Noah-MP [*Niu, et al.*, 2009]. Noah-MP offers multiple options for choosing the modeling of certain physical phenomena. In this study, the soil moisture factor for stomatal resistance is of "Noah type" [*Niu, et al.*, 2009] and the runoff scheme is from "SIMGM" [*Niu, et al.*, 2007]. The soil column is 2 meter deep, below which is an unconfined aquifer. In order to represent the characteristics of the structural soil over the model domain, the saturated hydraulic conductivity, which is determined by the soil texture data, is enlarged by factor of ten (through calibration). The soil hydrology of Noah (soil moisture) is run at an hourly time step and runoff data are produced every three hours.

Noah-MP calculates the amount of water that runs off on and below the land surface. This quantity is used to provide RAPID with the water inflow from outside of the river network. David et al. [2009] presented a coupling technique using a hydrologically enhanced version of the Noah LSM called Noah-distributed [*Gochis and Chen*, 2003] that allows physically-based modeling of the horizontal movement of surface and subsurface water from the land surface to a river reach. In interest of a simpler coupling scheme, the work of David et al. [2009] has been modified. In this study, a flux coupler between Noah and RAPID is developed using the catchments available in the NHDPlus dataset.

The NHDPlus catchments contributing runoff to each river reach were determined as part of the NHDPlus development using a digital elevation model and its associated flow accumulation and flow direction grids. These grids have a native resolution of 30 m. The map of catchments is available in NHDPlus in both gridded (at 30-m resolution)

and vector formats in a shapefile. Running a land surface model at a 30-m resolution is very resource demanding. Therefore, a coarser resolution of 900 m cell size is chosen. The shapefile of NHDPlus catchment boundaries is converted to a grid of size 900 m. Within this conversion process, the accuracy of the boundaries of the catchments is lowered but the catchment boundaries are reasonably respected and the computational cost of the land surface model calculations is reasonable. For each 3-hour output of the Noah model, surface and subsurface runoff data is superimposed onto the catchment grid, and all runoff that corresponds to the catchment of each river reach is summed and used as the water inflow to the river reach. Figure 13 shows the principle of the flux coupler in which the 900-m runoff data generated by the Noah model is superimposed to the 900-m map of NHDPlus catchment COMIDs to determine the lateral inflow for NHDPlus reaches used by RAPID.

Therefore, no horizontal routing is used between the land surface and the river network in the proposed scheme. This differs from some other models that use runoff from a one-dimensional model to force a river routing model. For instance, the two dimensional wave equation is used in Gochis and Chen [2003] or the linear reservoir equation is used in Ledoux et al. [1989].

The coupling method used here can be adapted to any land surface model that computes surface and subsurface runoff on a grid. This coupling technique is automated in a Fortran program.

3.4.3. Meteorological forcing

Land surface models need meteorological forcing in order to compute the water and the energy balance at the surface. The Noah LSM requires seven meteorological parameters: precipitation, specific humidity, air temperature, air pressure, wind speed, downward shortwave and downward longwave radiation. Hourly precipitation is

obtained from NEXRAD and downscaled from its original resolution (4.763 km) to 900 m using the method developed in Guan et al. [2009]. All other meteorological parameters are downloaded from the 3-hourly North American Regional Reanalysis (NARR) and converted from its original resolution (32.463 km) to 900 m using a simple triangle-base linear interpolation. All meteorological data are prepared for four years (01 January 2004 – 31 December 2007).

3.5. RESULTS

The framework for computation of river flow that is developed in the previous section is used to calculate river flow in all 5175 river reaches of the Guadalupe and San Antonio River Basins for four years (01 January 2004 – 31 December 2007). In this section, flow wave celerities in several subbasins are estimated from measurements, the model parameters used in RAPID are presented, and computed flows are compared to observed flows. Issues related to the time step used in RAPID and to the simulated wave celerities are also presented.

3.5.1. Estimation of wave celerities

The USGS Instantaneous Data Archive (http://ida.water.usgs.gov/ida/) provides 15-minute flow data that can be used to determine the flow wave celerity. Data at fourteen gaging stations within the two basins studied are obtained from IDA over two time periods (01 January 2004 – 30 June 2004 and for 01 January 2007 – 30 June 2007). The maximum lagged cross-correlation between hydrographs at two consecutive gaging stations is used to determine the flow wave celerity. Figure 14 shows the correlation as a function of increasing lag time between three different sets of consecutive gaging stations. The lag time giving the maximum correlation is taken as the travel time τ_{travel} for the flow wave between the two stations. The lag times are estimated for eleven sets of two stations and are shown on Table 1. Lag times of 0 s are reported at two stations, where the flow wave is probably too fast to be captured by 15-minute measurements. The wave celerity c is then computed using Equation (19)

$$c = \frac{d}{\tau_{travel}} \tag{19}$$

where d is the distance between two stations. The NHDPlus Flow Table Navigator Tool (http://www.horizon-systems.com/nhdplus/tools.php) is used to estimate the curvilinear distance between two stations along the NHDPlus river network that are shown on Table 1. The wave celerity has been estimated for eleven subbasins within the San Antonio and Guadalupe river basins. Table 2 shows the values that are obtained for the two time periods considered, as well as their average. Figure 15 shows the corresponding subbasins.

3.5.2. Parameters used in RAPID

RAPID needs two vectors of parameters $\mathbf{k}$ and $\mathbf{x}$ that can either be determined using physically-based equations, through optimization, or a combination of both. In this study, daily stream flow data are obtained from the USGS National Water Information System (http://waterdata.usgs.gov/nwis) in order to use the built-in parameter estimation. Within the Guadalupe and San Antonio river basins, NWIS has 74 gages that measure flow, 32 of them having full records of daily measurements the four years studied (01 January 2004 – 31 December 2007). These 32 stations are used for parameter estimation.

Four sets of model parameters – denoted by the superscripts α, β, γ and δ – are used in this study. These sets of parameters are all based on Equation (17) which is used with a uniform wave celerity of $c^0 = 1km \cdot h^{-1} = 0.28m \cdot s^{-1}$ throughout the basin or with the celerities c_j determined based on the IDA lagged cross-correlation study.

The first set, $(\mathbf{k}^{\alpha}, \mathbf{x}^{\alpha})$ is obtained from parameter estimation using the uniform wave celerity $c^0 = 0.28m \cdot s^{-1}$ and the resulting values of the two multiplying factors λ_k and λ_x of Equation (17) are:

$$k_j^{\alpha} = \lambda_k^{\alpha} \cdot \frac{L_j}{c^0} \quad , \quad x_j^{\alpha} = \lambda_x^{\alpha} \cdot 0.1$$
$$\lambda_k^{\alpha} = 0.125 \quad , \quad \lambda_x^{\alpha} = 0.9375 \tag{20}$$

The parameters $(\mathbf{k}^{\beta}, \mathbf{x}^{\beta})$ are determined without optimization using the celerities c_j determined based on the IDA lagged cross-correlation study and set to:

$$k_j^{\beta} = \lambda_k^{\beta} \cdot \frac{L_j}{c_j} \quad , \quad x_j^{\beta} = \lambda_x^{\beta} \cdot 0.1$$
$$\lambda_k^{\beta} = 1 \quad , \quad \lambda_x^{\beta} = 1 \tag{21}$$

The third set of parameters $(\mathbf{k}^{\gamma}, \mathbf{x}^{\gamma})$ is obtained through optimization using the celerities c_j determined based on the IDA lagged cross-correlation study and the resulting values are:

$$k_j^{\gamma} = \lambda_k^{\gamma} \cdot \frac{L_j}{c_j} \quad , \quad x_j^{\gamma} = \lambda_x^{\gamma} \cdot 0.1$$
$$\lambda_k^{\gamma} = 0.75 \quad , \quad \lambda_x^{\gamma} = 1.625 \tag{22}$$

The optimization converges to a value of $\mathbf{k}$ that is 25% smaller than that estimated with the IDA lagged cross-correlation, suggesting that a faster flow wave in the river network produces better flow calculations. In the present study, routing on the land surface from the catchment to its corresponding reach is not modeled. Therefore, one would expect that the optimized flow celerity in the river network would be slower than

that estimated from river flow observations, which is not the case here. This suggests that runoff is produced too fast which could be because runoff in land surface models is often calibrated based on a lumped value at the downstream gage of a basin.

The fourth set of parameters $(\mathbf{k}^{\delta}, \mathbf{x}^{\delta})$ is determined for a better match of celerity calculations, as explained later in this paper.

3.5.3. Time step of RAPID simulation

Cunge [1969] showed that the Muskingum method is stable for any $x \in [0, 0.5]$ and that the wave celerity computed by the Muskingum method approaches the theoretical wave celerity of the kinematic wave equation if the time step of the river routing equals the travel time of the wave (for $x = 0.5$), as shown in Equation (23):

$$\forall j \in [1, m] \qquad c_j \simeq \frac{L_j}{\Delta t} \tag{23}$$

However, both the celerity of flow and the length river reaches vary along the network; and the model formulation of RAPID allows for only one unique value of the time step Δt be chosen. In the Guadalupe and San Antonio River Basins, the mean length is 3 km and the median length is 2.4 km. The probability density function and the cumulative density functions for the lengths of river reaches are shown in Figure 16. The celerities estimated earlier are on the order of $c = 2.5 m \cdot s^{-1}$. Using the median value of the reach length along with $c = 2.5 m \cdot s^{-1}$, Equation (23) gives $\Delta t = 960 s$. In order to have an integer conversion between the river routing time step and the land surface model time step (3 hours), a value of $\Delta t = 900 s = 15 \min$ is chosen.

3.5.4. Analysis of the quality of river flow computation

For various model simulations, the average and the root mean square error (RMSE) of computed flow rate are calculated using daily data and are given in Table 3. The Nash efficiency [*Nash and Sutcliffe*, 1970] is bounded by the interval $[-\infty,1]$ and gives an estimate of the quality of modeled river flow computations when compared to observations. An efficiency of 1 corresponds to a perfect model and 0 corresponds to a model producing the mean of observations. The Nash efficiency is also given in Table 3. The results shown for a lumped model correspond to when runoff from Noah is accumulated at the gage directly without any routing. The average values of flow in RAPID simulations are tied to the amount of runoff water calculated by the Noah LSM and the bias generated by the land surface model cannot be fixed by RAPID. However, the internal connectivity of the NHDPlus river network is well translated in RAPID and continuity within RAPID is verified since the flow rates in the lumped simulation and in all four simulations of RAPID are the same.

Furthermore, all RAPID simulations (regardless of what parameters are used) lead to a smaller RMSE and a higher Nash Efficiency than the lumped runoff. This shows that an explicit river routing scheme allows obtaining better stream flow calculations than a simple lumped runoff scheme, as expected.

Within the different RAPID simulations, the set of parameters $(\mathbf{k}^{\gamma},\mathbf{x}^{\gamma})$ gives the best results for RMSE and Nash efficiency, followed by $(\mathbf{k}^{\beta},\mathbf{x}^{\beta})$, $(\mathbf{k}^{\alpha},\mathbf{x}^{\alpha})$ and $(\mathbf{k}^{\delta},\mathbf{x}^{\delta})$. Therefore, a greater spatial variability in the values of k contributes to the quality of model outputs, and the built-in optimization in RAPID allows achieving greater quality in model results. An example hydrograph for the Guadalupe River near Victoria TX is shown in Figure 17, and is computed using $(\mathbf{k}^{\gamma},\mathbf{x}^{\gamma})$.

3.5.5. Comparison between estimated and computed wave celerities

In order to assess the capacity of the modeling framework to reproduce momentum, the celerity of the flow wave in outputs from RAPID are computed. Fifteen-minute river flow is computed with RAPID, and the lagged cross-correlation presented earlier is used to calculate the wave celerity within the RAPID simulation. Table 2 shows the celerities that are computed from RAPID outputs. In the first three sets of model parameters used, the wave celerities simulated in RAPID are greater than those observed. One can also notice than even for $(\mathbf{k}^{\beta}, \mathbf{x}^{\beta})$, the model-simulated celerities are different than the observed celerities which are used to determine the vector $\mathbf{k}^{\beta}$ itself. This was predicted by Cunge [1969] who showed that the difference between the celerity of the kinematic wave equation and that computed using the Muskingum method is a function of both x and the quotient $\frac{\Delta t}{L_j}$. Only the specific values $x = 0.5$ and $\frac{\Delta t}{L_j} = 1$ allow obtaining the same celerity. Furthermore, the work herein is done in a river network, and the celerity estimated between two points does not correspond only to the main river stem but rather to a combination of all river reaches present in the network in between the two points. The ratio of the average celerities from RAPID using $(\mathbf{k}^{\beta}, \mathbf{x}^{\beta})$ over the average observed celerities is 1.44. As a final experiment, a new set of parameters $(\mathbf{k}^{\delta}, \mathbf{x}^{\delta})$ is created to account for the faster waves in RAPID.

$$\begin{aligned} k_j^{\delta} &= \lambda_k^{\delta} \cdot \frac{L_j}{c_j} \quad , \quad x_j^{\delta} = \lambda_x^{\delta} \cdot 0.1 \\ \lambda_k^{\delta} &= 1.44 \quad , \quad \lambda_x^{\delta} = 1 \end{aligned} \tag{24}$$

Table 2 shows that the parameters $(\mathbf{k}^{\delta}, \mathbf{x}^{\delta})$ allow for wave celerities that are closer to the observed ones than the celerities obtained with the other sets of parameters. The average flow wave celerity over the 11 calculations in RAPID is within 3% of that estimated with IDA flows. Unfortunately, these closer wave celerities also lead to a decrease in the quality of RMSE and Nash Efficiency. Therefore, model celerities closer to celerities estimated from observations can be obtained, but generally deteriorate other statistics of calculations. Again, this might be due to runoff produced too fast.

3.5.6. Scalability of parallel computation

The work presented here focuses on a river network with 5,175 river and water body reaches. However, all the tools and datasets used are available for the contiguous United States. The river network of the NHDPlus dataset has about 3 million reaches for the United States. Adapting the proposed framework to simultaneously compute flow and volume of water in all mapped water bodies of the contiguous United States require solving matrix equations of size 3 million. For such a large scientific problem, parallel computing is needed and scalability of parallel computations necessary.

Through the use of mathematical and optimization libraries that run in a parallel computing environment, RAPID can be applied on several processors. Figure 18 shows basic tests that are performed on the Lonestar supercomputer in order to assess the scalability of the river network model. Five simulations are run, with increasing numbers of processors used. The same model simulation of river flow in the Guadalupe and San Antonio river basins, over 4 years, at a 900-second time step is used for all tests reported in Figure 18; but the number of processors used for computation differs. Three computing times are given: the total central processing unit (CPU) time is summation of the CPU time of all the processors. The total CPU time is divided by the number of processors to give the average CPU time. Finally, the wall-clock time is the time

between the start of all computations and the end of the last calculation of the last processor. Figure 18 shows that the duration of the simulation decreases with increasing number of processors up to 4 processors, and increases beyond 4 processors. Therefore, RAPID shows some scalability for the simulation considered. However, the problem size and the computing time do not yet justify the use of high performance parallel computing. The application of RAPID on a larger problem and the investigation of its performance and scalability are needed before considering its use in a parallel computing environment. In particular, the way inputs and outputs are handled in a parallel computing environment and various options for repartition of computations to reflect the connectivity of the river network should be investigated. Furthermore, river flow is a causal physical phenomenon by essence. The flows in river reaches that belong to the same river network are connected and cannot be computed independently. In a classical time stepping environment such as the one developed in this study (i.e. a computing scheme in which all the spatial elements at the current time are calculated based on some or all of the spatial elements at the previous and current time), one cannot obtain results at the downstream-most point prior to having calculated everywhere else. Therefore, one way to leverage parallel computing is to use different processors on completely disconnected basins. Hence, the calculation at one node would be completely independent from the calculation at another node. Another option would be translating the problem in another time-space environment (such as a wave front transformation) where calculations would be disassociated.

3.6. CONCLUSIONS

NHDPlus is a GIS dataset that describes the networks of mapped rivers and water bodies of the United States. One of the main advantages of NHDPlus is that connectivity information for the river networks is available. Therefore, this dataset offers possibilities for the development of river routing models that simultaneously calculate flow and volume of water in all water bodies of the nation. The research presented in this paper investigates how to develop a river network model, and how NHDPlus can serve as its river network. All tools and datasets used are available for the contiguous United States, but this research addresses a much smaller area: the Guadalupe and San Antonio River Basins in Texas. Graph theory is applied to a river network to create a network matrix that is used to develop a vector-matrix version of the Muskingum method and applied to a river network in a model called RAPID. It has been shown that a GIS-based hydrographic dataset can be used as the river network for a river model to compute flow in large networks of thousands of reaches, including ungaged locations. A simple flux coupler for connecting a land surface model with an NHDPlus river network is presented. No horizontal routing of flow from the land surface to the river network is used in this study, and such an addition would help improve model calculations. An inverse method is developed to estimate model parameters in RAPID using available gage measurements. Wave celerities are estimated in several locations of the basin studied. RMSE and Nash efficiency of computed flow rate in four RAPID simulations are compared with a basic lumped model where runoff is directly accumulated at the gage, with gage measurements and among themselves. RAPID produces better RMSE and Nash efficiency than the lumped model. Although the quality of RAPID calculations is tied to the quantity of runoff generated by the land surface model that provides runoff, continuity within

RAPID is verified since the average flow rate is conserved. Spatial variability of parameters enhances the RMSE and Nash efficiency of RAPID calculations. Wave celerities are reproduced within a few percents with the model proposed, although wave celerities closer to those estimated from gage data generally deteriorate the other statistics of calculations. This might be due to runoff being produced too fast. The parameters used in this study are simple, but could be improved based on information available in NHDPlus such as slope, mean flow and velocity of all reaches. The matrix formulation in RAPID can be transferred in a parallel computing environment. However, performance issues have barely been tackled in this study, and a thorough general profiling or RAPID in a parallel computing environment would help improve the model. In particular, the investigation of the appropriate handling of large amounts of inputs and outputs is important. Such work is needed before using RAPID in a parallel computing environment to address larger scales in the United States, like at the state or federal level which would represent a square matrix of size 3 million.

Table 1 Lag time (s) estimated using the lagged cross-correlation in the Guadalupe and San Antonio River Basins, both from IDA measurements and from RAPID model runs; and distance (km) between gaging stations

		Location of the two consecutive streamflow gages										
		Ingram - Kerrville	Kerrville - Comfort	Comfort - Spring Branch	Sattler - Gonzales	Gonzales - Cuero	Cuero - Victoria	Schroeder - Victoria	Bandera - Macdona	Macdona - Elmendorf	Elmendorf - Falls City	Falls City - Goliad
Lag time (s) from IDA	2004	7200	18900	60300	163350	132300	69750	*0*	20700	*0*	126000	162000
	2007	6300	18900	59400	130050	108450	69750	8100	37350	15300	91800	126000
	average	6750	18900	59850	146700	120375	69750	4050	29025	7650	108900	144000
Lag time (s) from RAPID outputs	RAPID (k^{α},x^{α})	7200	8100	31950	88650	27000	36450	4500	47700	25650	19800	47700
	RAPID (k^{β},x^{β})	6300	10350	46800	130500	84150	32850	4050	24300	10800	81450	121950
	RAPID (k^{γ},x^{γ})	4500	8100	35100	103500	71100	33300	3600	18000	9000	62100	95400
	RAPID (k^{δ},x^{δ})	8100	13500	67050	174600	110250	68400	4500	35100	17100	117900	168300
Distance (km)		17.30	40.40	100.85	203.50	110.72	100.71	23.44	116.49	63.49	87.73	137.28

Table 2 Wave celerities (m/s) estimated using the lagged cross-correlation in the Guadalupe and San Antonio River Basins, both from IDA measurements and from RAPID model runs

Wave celerity (m/s)		Location of the two consecutive streamflow gages										
		Ingram - Kerrville	Kerrville - Comfort	Comfort - Spring Branch	Sattler - Gonzales	Gonzales - Cuero	Cuero - Victoria	Schroeder - Victoria	Bandera - Macdona	Macdona - Elmendorf	Elmendorf - Falls City	Falls City - Goliad
from IDA	2004	2.40	2.14	1.67	1.25	0.84	1.44	∞	5.63	∞	0.70	0.85
	2007	2.75	2.14	1.70	1.56	1.02	1.44	2.89	3.12	4.15	0.96	1.09
	average	2.57	2.14	1.69	1.41	0.93	1.44	2.89	4.37	4.15	0.83	0.97
from RAPID outputs	RAPID (k^{α},x^{α})	2.40	4.99	3.16	2.30	4.10	2.76	5.21	2.44	2.48	4.43	2.88
	RAPID (k^{β},x^{β})	2.75	3.90	2.16	1.56	1.32	3.07	5.79	4.79	5.88	1.08	1.13
	RAPID (k^{γ},x^{γ})	3.84	4.99	2.87	1.97	1.56	3.02	6.51	6.47	7.05	1.41	1.44
	RAPID (k^{δ},x^{δ})	2.14	2.99	1.50	1.17	1.00	1.47	5.21	3.32	3.71	0.74	0.82

Table 3 Comparison of observed and simulated flows at four locations within the Guadalupe and San Antonio River Basins

	Average daily stream flow (m3/s)					
Gaging station	Observed	Lumped	RAPID (k^{α},x^{α})	RAPID (k^{β},x^{β})	RAPID (k^{γ},x^{γ})	RAPID (k^{δ},x^{δ})
Guadalupe River near Victoria	80.97	61.98	62.00	62.00	61.99	62.02
Guadalupe River at Sattler	22.04	6.62	6.63	6.63	6.62	6.63
Coleto Creek near Victoria	3.99	13.74	13.74	13.74	13.74	13.74
San Antonio River at Goliad	37.54	34.97	34.99	34.99	34.99	35.00
Mean	36.14	29.33	29.34	29.34	29.34	29.35
	RMS error (m3/s) using daily averages					
Gaging station	Observed	Lumped	RAPID (k^{α},x^{α})	RAPID (k^{β},x^{β})	RAPID (k^{γ},x^{γ})	RAPID (k^{δ},x^{δ})
Guadalupe River near Victoria	N/A	96.63	80.21	77.38	75.08	94.47
Guadalupe River at Sattler	N/A	39.38	39.26	39.22	39.27	39.16
Coleto Creek near Victoria	N/A	25.66	26.34	26.24	26.13	26.38
San Antonio River at Goliad	N/A	41.82	42.30	44.64	43.24	49.14
Mean	N/A	50.87	47.03	46.87	45.93	52.29
	Nash efficiency using daily averages					
Gaging station	Observed	Lumped	RAPID (k^{α},x^{α})	RAPID (k^{β},x^{β})	RAPID (k^{γ},x^{γ})	RAPID (k^{δ},x^{δ})
Guadalupe River near Victoria	N/A	0.52	0.67	0.68	0.71	0.54
Guadalupe River at Sattler	N/A	-0.02	-0.01	-0.01	-0.01	-0.01
Coleto Creek near Victoria	N/A	-0.25	-0.31	-0.30	-0.29	-0.32
San Antonio River at Goliad	N/A	0.57	0.56	0.51	0.54	0.41
Mean	N/A	0.21	0.23	0.22	0.24	0.16

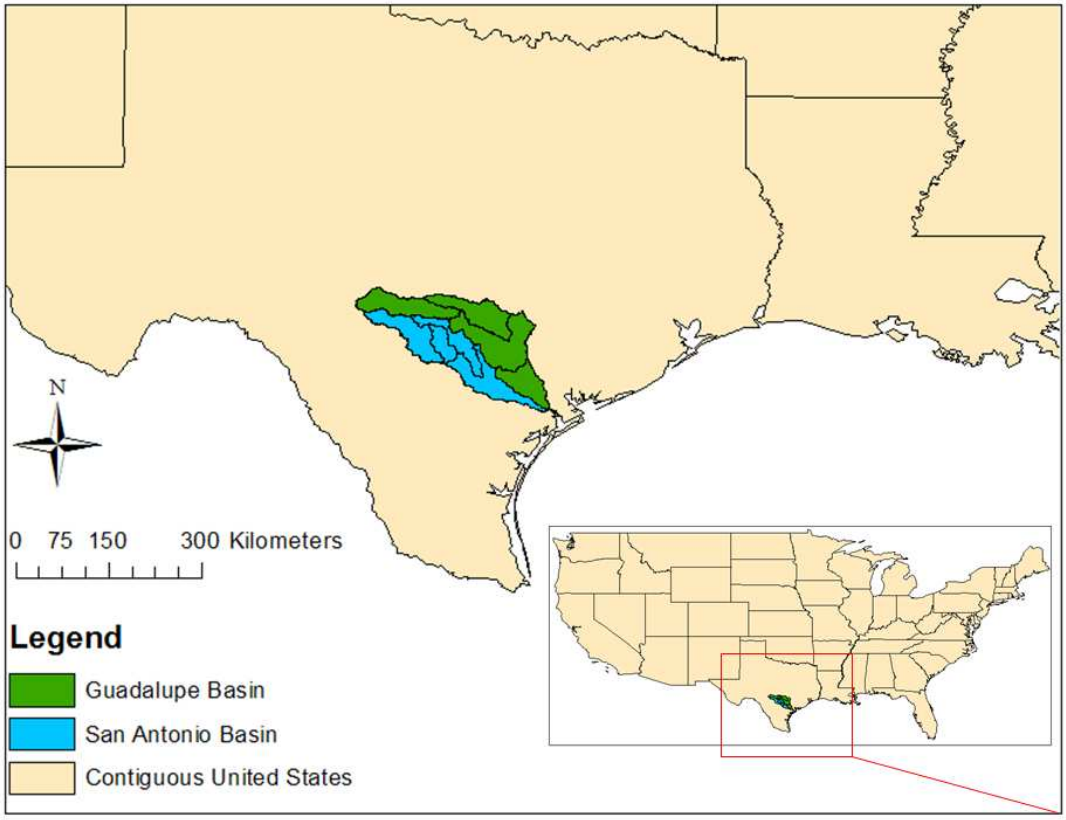

Figure 9 Guadalupe and San Antonio Basins

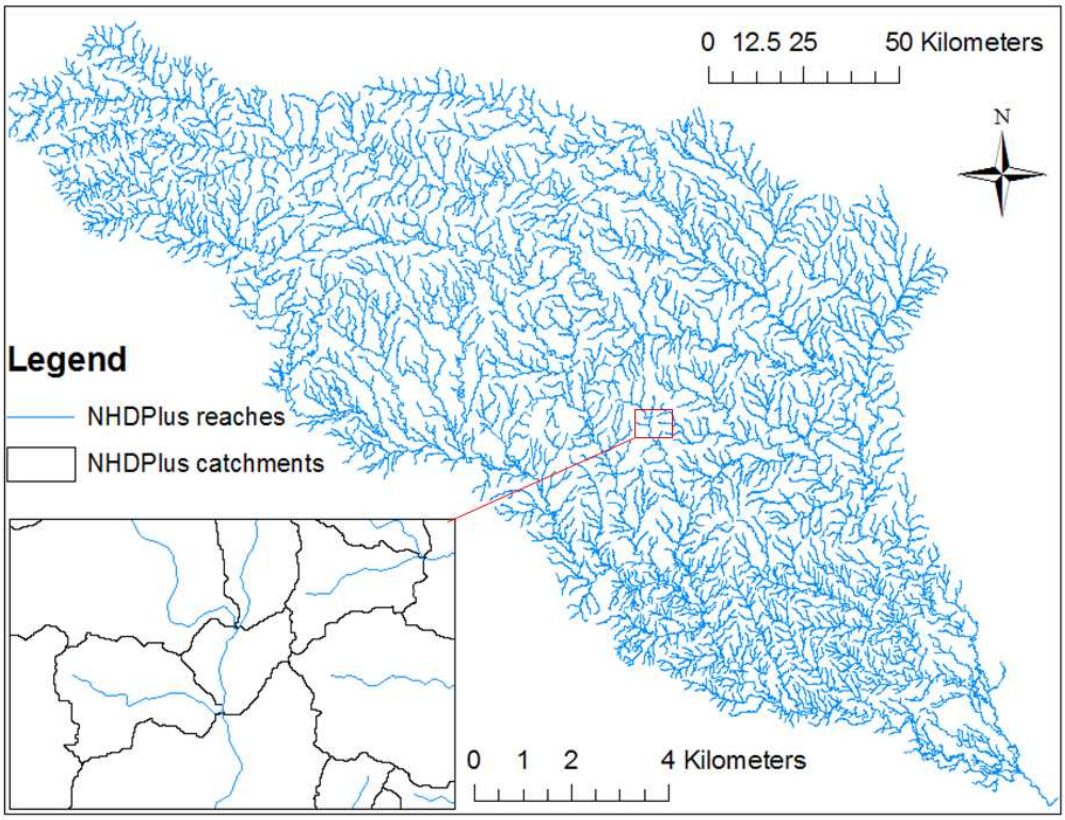

Figure 10 NHDPlus river network and catchments for the Guadalupe and San Antonio Basins

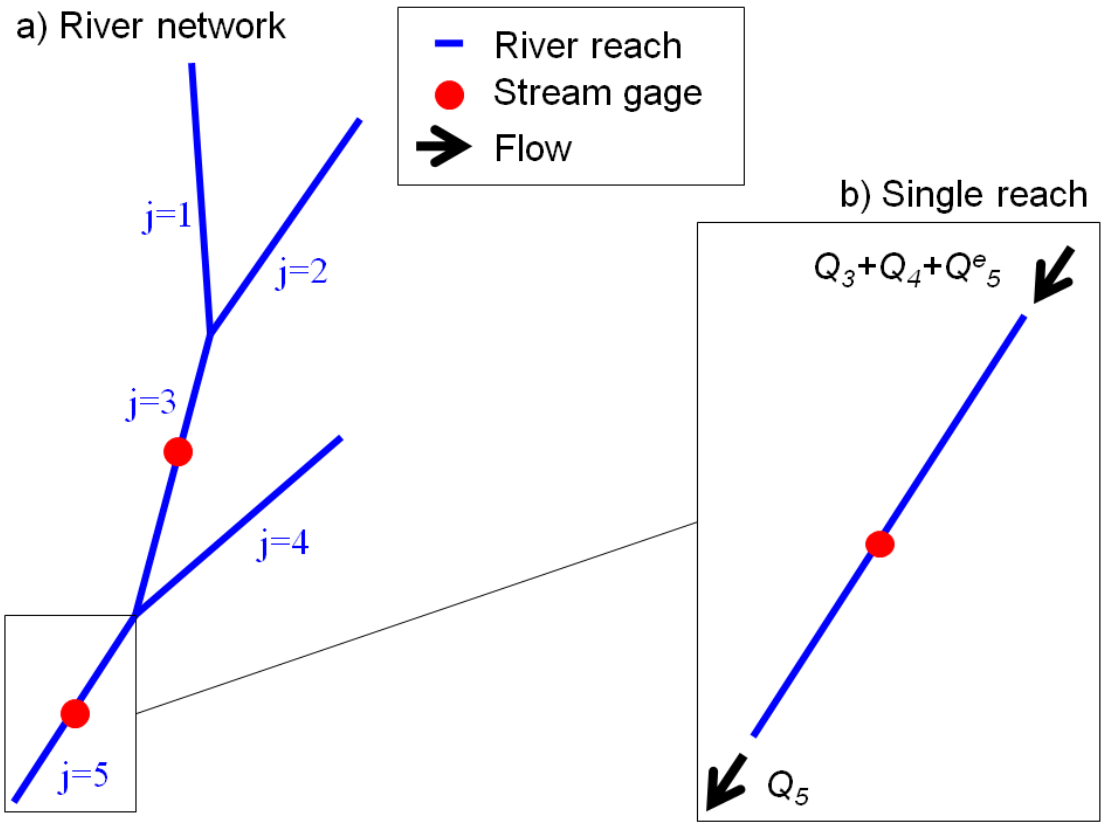

Figure 11 River network

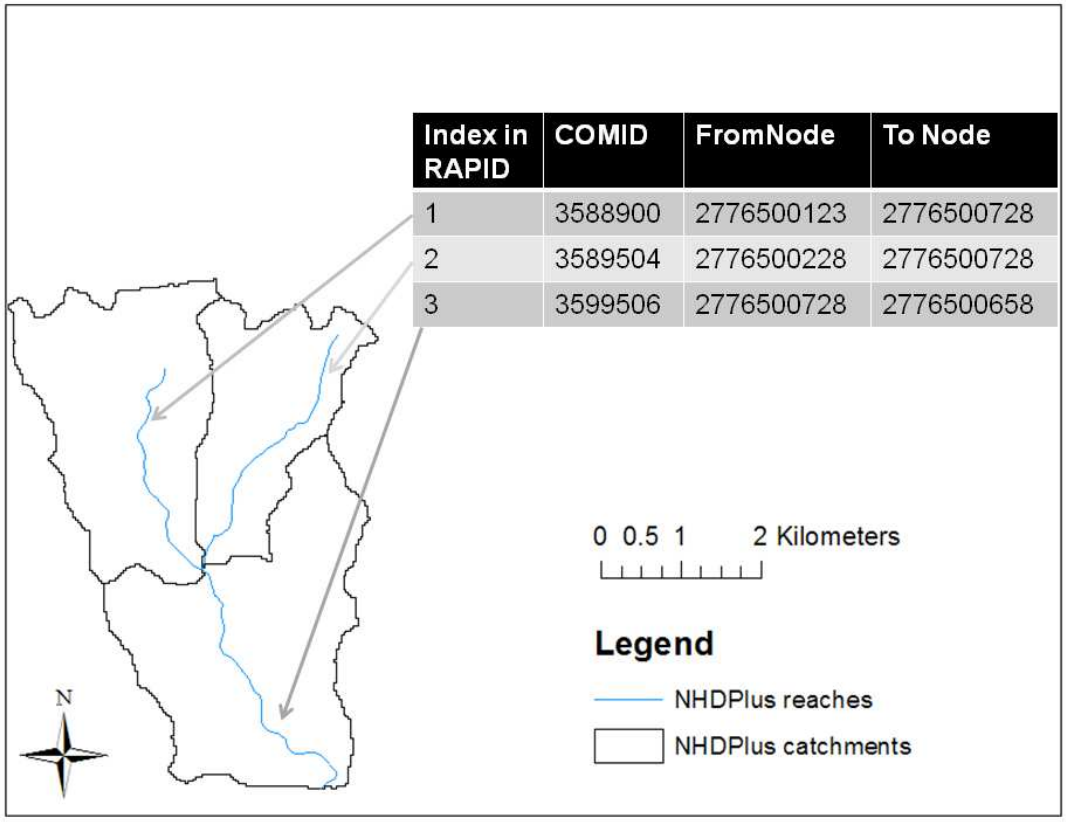

Figure 12 NHDPlus connectivity between reaches, nodes and catchments

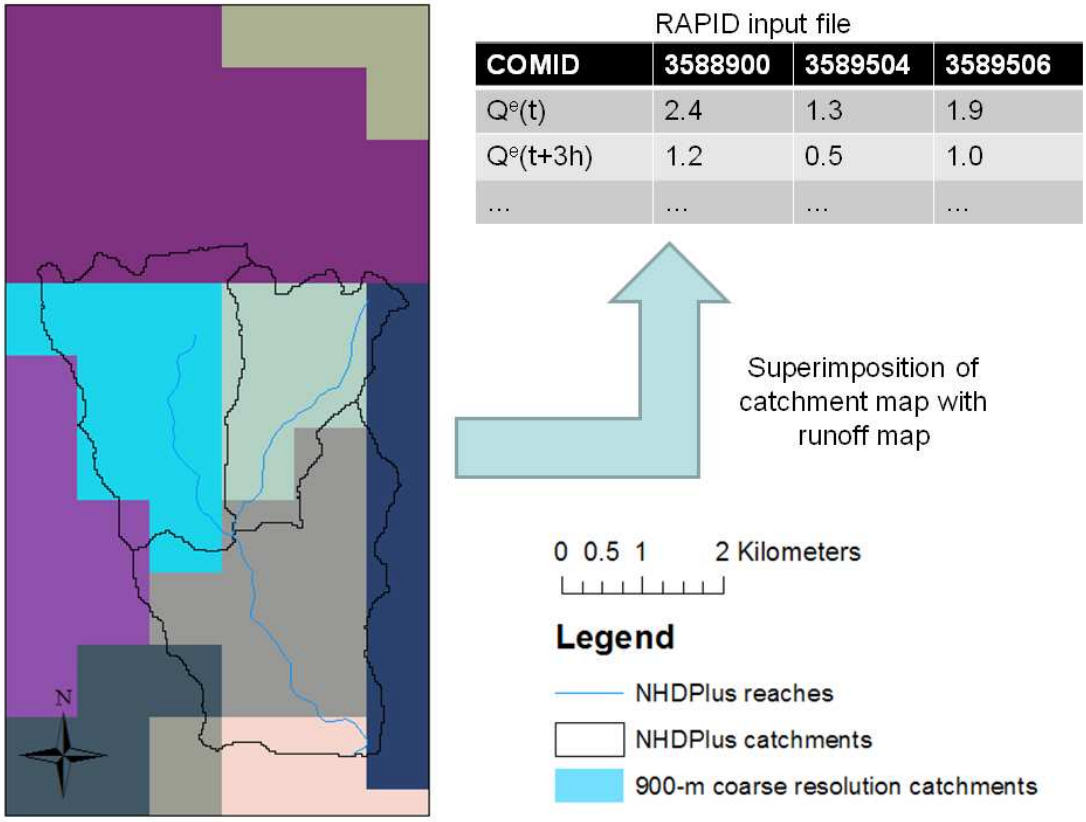

Figure 13 Principle of flux coupler between Noah and RAPID

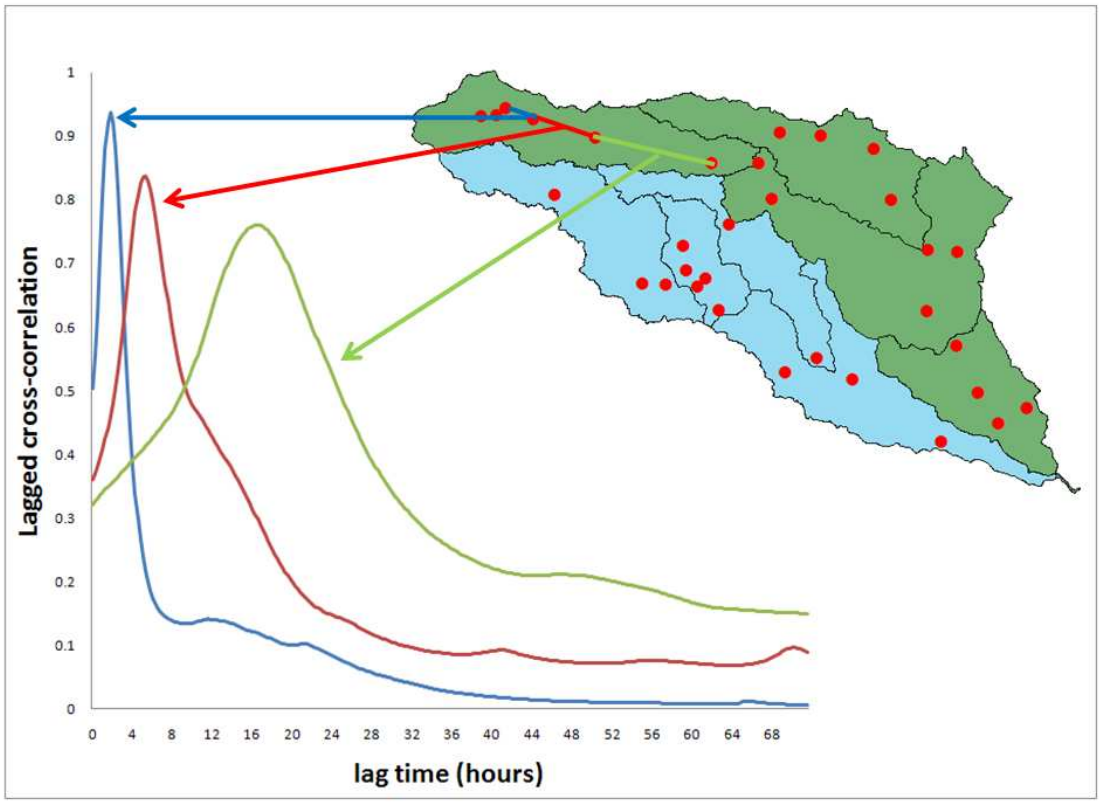

Figure 14 Lagged cross-correlation as a function of lag time

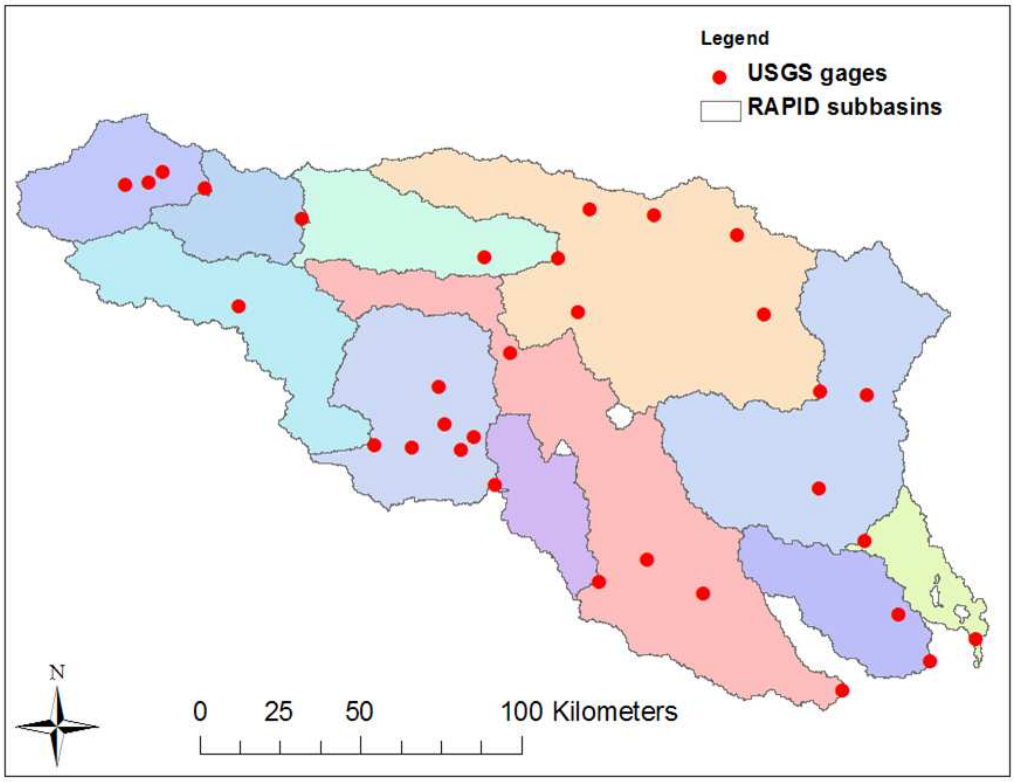

Figure 15 Wave celerities are estimated for eleven different subbasins within the Guadalupe and San Antonio river basins. The same subbasins are used for distributed parameters in RAPID

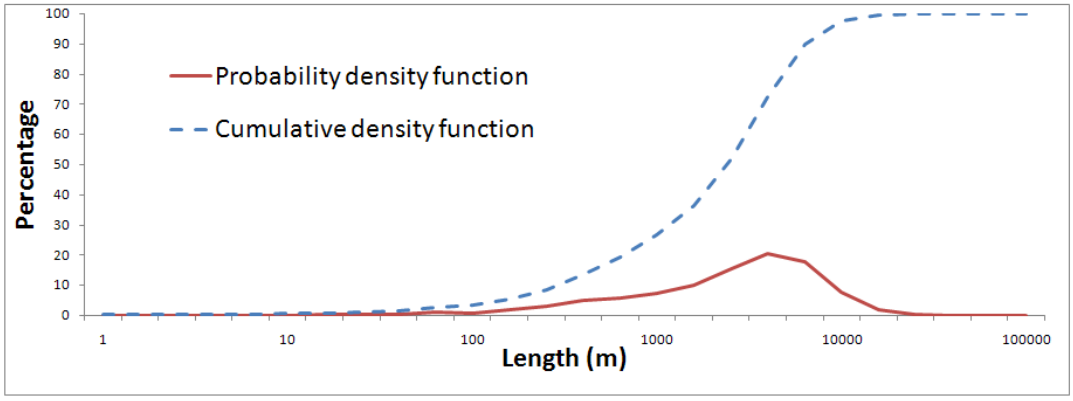

Figure 16 Statistics of river reach lengths in Guadalupe and San Antonio River Basins

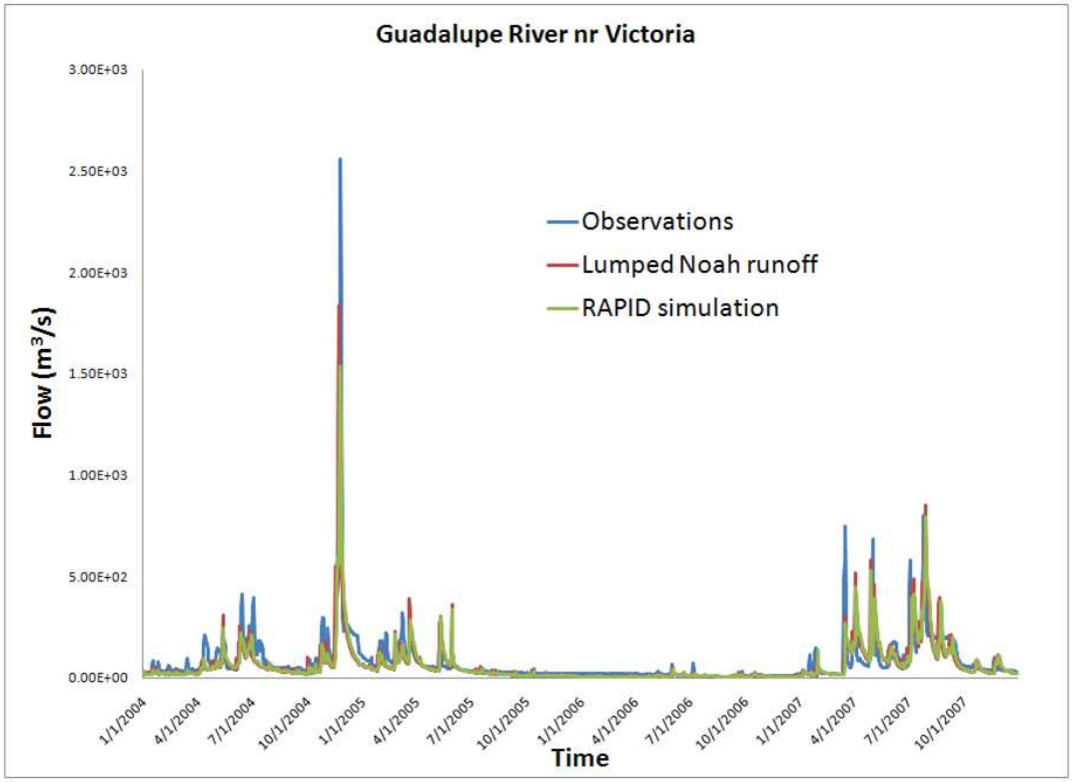

Figure 17 Hydrograph of observed, lumped and routed flows for the Guadalupe River near Victoria, using (k^{γ},x^{γ})

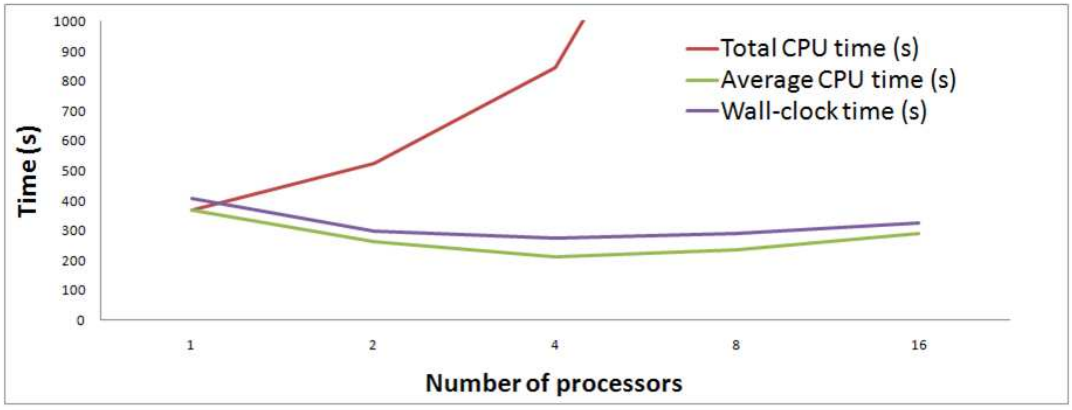

Figure 18 Scalability of RAPID computations

Chapter 4: Using RAPID with the SIM-France hydro-meteorological model

4.1. ABSTRACT

SIM-France is a large connected atmosphere/land surface/groundwater modeling system that simulates the water cycle throughout metropolitan France. The work presented in this study investigates the replacement of the river routing scheme in SIM-France by a river network model called RAPID to improve the capacity to relate simulated flows to river gages and to provide stable computation of river/aquifer exchanges. RAPID was coupled with SIM-France over a ten-year period and results compared with those of the previous river routing scheme. We found that while the formulation of RAPID improved the functionality of SIM-France, the flow simulations are comparable in accuracy to those previously obtained by SIM-France. Sub-basin parameterization was found to improve model results. A single criterion for quantifying the quality of river flow simulations using several river gages globally in a river network is developed that normalizes the square error of modeled flow to allow equal treatment of all gaging stations regardless of the magnitude of flow. The use of this criterion as the cost function for parameter estimation in RAPID allows better results than by increasing the degree of spatial variability in model parameters.

4.2. INTRODUCTION

In the past two decades, several large scale river routing schemes have been used along with land surface models for hydrologic modeling. Among the most notable applications of large scale river routing are TRIP [*Ngo-Duc, et al.*, 2007; *Oki and Sud*, 1998], the routing model of Lohmann et al. [*Lohmann, et al.*, 1996; 1998a; 1998b; 1998c; 2004; *Maurer, et al.*, 2001], that of Wetzel [*Abdulla, et al.*, 1996; *Nijssen, et al.*, 1997; *Wetzel*, 1994] and that of Olivera et al. [2000]. These approaches have been used along with land surface parameterization schemes to calculate river flow from runoff at the regional, continental and the global scale. MODCOU [*Ledoux, et al.*, 1989] is another model with routing capabilities that differs from the previously cited models in that it has separate horizontal routing of water for land surface and within the river system. MODCOU simulates flows throughout Metropolitan France (mainland France and Corsica) as part of the SIM-France modeling framework [*Habets, et al.*, 2008].

SIM-France is a large connected atmosphere, surface and groundwater model (see Figure 19) that involves coupling the national-scale atmospheric analysis system SAFRAN [*Durand, et al.*, 1993; *Quintana-Segui, et al.*, 2008], with the ISBA land surface model [*Boone, et al.*, 1999; *Noilhan and Planton*, 1989], and with the MODCOU hydrogeological model [*Ledoux, et al.*, 1989]. ISBA computes the vertical water and energy balance between the land surface and the atmosphere. The improved physics of the land surface parameterization of ISBA developed in Quintana-Seguí et al. [2009] are used in this study. Surface runoff and deep-soil drainage are computed by ISBA and transferred to MODCOU which computes the horizontal flow routing on the land surface, in rivers and in aquifers. Aquifers in MODCOU are modeled within the two main river

basins of France, the Seine and the Rhône, which together represent 30% of the land area of France.

MODCOU handles the calculations of flow and volume of water within the river network of SIM-France. This river network is made up of grid cells divided into a quad-tree pattern and the calculations of MODCOU are made for groups of quad-tree cells. Using groups of cells for calculations is advantageous for reducing computational costs but it limits the modularity of MODCOU. In particular, the location and number of gaging stations are difficult to modify, and bi-directional river/aquifer interactions on groups of cells are unstable.

The work presented herein investigates the impact of replacing the routing module used in MODCOU by a river network model called RAPID [see Chapter 3]. RAPID uses a matrix-based version of the Muskingum method to calculate flow and volume of water for each reach of a river network. RAPID was previously applied to a GIS vector river network in Chapter 3, and the present study shows how it can also be applied to a quad-tree gridded river network.

In this paper, the original river routing of MODCOU as well as that of RAPID are briefly presented followed by a ten-year application (1995-2005) of SIM-France comparing the two river routing applications.

4.3. MODELING FRAMEWORK

4.3.1. River modeling in SIM-France

The computational domain of SIM-France includes all of Metropolitan France, including Corsica. Parts of Spain, Switzerland, Germany and Belgium are also included where their drainage area flows through France, as shown in Figure 20. The total surface area of the computational domain is 610,000 km^2.

Surface routing and river routing in SIM-France are done by MODCOU [*Ledoux, et al.*, 1989]. The surface and river networks of SIM-France and their connectivity were created using a routine called HydroDem [*Leblois and Sauquet*, 2000] and consist of 193,861 surface cells and 24,264 river cells, each river cell being a particular surface cell. The surface area covered by the river cells is 65,000 km^2. The surface network uses a quad-tree structure with cell sizes of 1 km, 2 km, 4 km and 8 km. The river network has cell sizes of 1 km and 2 km. The smaller quad-tree cells are used at the conference of branches of the river network for better representation of the network connectivity and at basin boundaries for more accurate basin surface area.

The connectivity between river cells is given by a table that provides for each downstream river cell up to four upstream river cells. There are no loops or divergences in the river network of SIM-France. The connectivity between catchments and rivers is given by a table that provides for each surface cell a unique downstream cell where its runoff enters the river.

For both surface and river routing, the calculations of flow and volume of water within MODCOU are carried out using groups of cells as computing elements, therefore minimizing the amount of calculations compared to computing for all cells separately. These groups of cells – or isochrone zones – are based on the notion of isochronism

developed by Villeneuve and Leblanc [1978]. An isochrone is a line representing a constant time of travel to a reference point downstream. An isochrone zone is the area between two successive isochrones. This zone is represented by a set of cells which are a single computational unit in MODCOU. Both the land surface isochrones and river isochrones of MODCOU have three-hour time intervals, which means that the time of travel between the upstream-most and the downstream-most cell in a given isochrone zone is approximately three hours. All the isochrones of a given network are determined using the travel time between connected cells which is estimated based on topography and on the geometry of the quad-tree mesh. For surface cells and river cells, the travel time $\tau_{i,j}$ between two consecutive cells i and j is calculated using the distance $d_{i,j}$ between the two cells and the slope $s_{i,j}$, as shown in Equation (25):

$$\tau_{i,j} = \alpha \cdot \frac{d_{i,j}}{\sqrt{s_{i,j}}} \tag{25}$$

where a unique value of α is calibrated for each major basin.

Figure 21 shows an example of the isochrone zones and connectivity between surface cells and river cells in MODCOU for the Ardèche River Basin. Figure 21a) shows the Ardèche River, its basin and three river gages. Figure 21b) shows the river isochrone zones of the Ardèche River. Figure 21c) shows the surface isochrone zones corresponding to the upstream-most river isochrone zone. Each surface cell belongs to a surface isochrone zone, but only the isochrone zones corresponding to one river isochrone zone are shown of Figure 21c) for clarity. The units used for isochrone zones are the number or MODCOU 3-hour time steps to the outlet (here the Mediterranean).

The quad-tree structure of increasing resolution can be seen at the boundary of the basin in Figure 21c).

In MODCOU, the volume of water V^{out} that discharges across each isochrone line in a computation time step is calculated differently for the surface network and for the river network. For routing on the land surface, all the volume of water V available in the isochrone zone is transferred to the downstream zone, as shown in Equation (26):

$$V^{out} = V \tag{26}$$

For routing in the river network, V^{out} is proportional to the volume of water V available within the isochrone zone as shown in Equation (27):

$$V^{out} = \beta \cdot V \tag{27}$$

where $\beta \in [0,1]$ is manually calibrated and usually set constant for large basins. Equation (27) can be viewed as the linear reservoir equation associated with a first-order explicit development of the continuity equation. The variation of volume related to lateral inflow and groundwater inflow of water are added to the volume V before calculating V^{out}. In SIM-France, β has four possible values: 0.5, 0.7, 0.8 and 0.9 as shown in Figure 22.

Equation (27) is applied to isochrone zones. Hence, the volume of water within each isochrone zone needs be partitioned among its several river cells before computation of the river-aquifer exchanges. This interaction depends on the aquifer head, on the river

head – assumed constant – and on the volume of water in the river cell when the river infiltrates water into the aquifer. The partitioning of water volume among all cells of an isochrone zone is done using a weighted average of the total amount of water.

This formulation has several inconsistencies, especially when the junction between two streams lies in the interior of an isochrone zone. This can have a consequence in the river-aquifer interaction, but also in the computation of the river flow. Furthermore, using only one set of isochrones in each basin can lead to two gages being located in one isochrone zone, in which case the flow computed by MODCOU has to match the flow at two different gaging stations. In order to avoid such inconsistencies, MODCOU uses a unique set of isochrone zones for each gage, such that each gage is the downstream-most river cell in its isochrone zone. Therefore, several flow calculations can be performed for a given cell, if the given cell belongs to several isochrone zones, which is inefficient and requires time consuming processing work in case of change of number or locations or river gages. The work done herein aims at simplifying the river modeling done within SIM-France.

4.3.2. RAPID

RAPID [see Chapter 3] is a river network model that uses a matrix-based version of the Muskingum routing scheme to calculate discharge simultaneously through a river network. RAPID was first applied to the Guadalupe and San Antonio River Basins in Texas using a vector-based river network extracted from a geographic information system dataset called NHDPlus [*USEPA and USGS*, 2007]. The governing equation used in RAPID is the following:

$$\left(\mathbf{I}-\mathbf{C}_1\cdot\mathbf{N}\right)\cdot\mathbf{Q}\left(t+\Delta t\right)=\mathbf{C}_1\cdot\mathbf{Q}^{\mathbf{e}}\left(t\right)+\mathbf{C}_2\cdot\left(\mathbf{N}\cdot\mathbf{Q}\left(t\right)+\mathbf{Q}^{\mathbf{e}}\left(t\right)\right)+\mathbf{C}_3\cdot\mathbf{Q}\left(t\right) \qquad (28)$$

where t is time and Δt is the river routing time step. The bolded notation is used for vectors and matrices. All matrices are square. $\mathbf{I}$ is the identity matrix. $\mathbf{N}$ is the river network connectivity matrix which has a value of one in element $N_{i,j}$ if reach j flows into reach i and zero elsewhere. $\mathbf{C_1}, \mathbf{C_2}$ and $\mathbf{C_3}$ are parameter matrices which depend on Muskingum k, x and time step Δt. $\mathbf{Q}(t)$ is a vector of outflows from river reaches, and $\mathbf{Q^e}(t)$ is a vector of lateral inflows to these reaches from land surface runoff or groundwater inflow. The number of river quad-tree cells – here 24,264 – is used for dimension of all vectors and matrices, each element of the vectors corresponding to one river cell.

Provided with a vector of lateral inflows $\mathbf{Q^e}(t)$, RAPID calculates the flow and volume of water in all reaches of a river network, therefore allowing coupling of a river network to most land surface models and groundwater models. A different value for the parameters k and x of the Muskingum method can be assigned for each river quad-tree cell, and RAPID uses two vectors $\mathbf{k}$ and $\mathbf{x}$ as input which are used to compute the values of the matrices $\mathbf{C_1}, \mathbf{C_2}$ and $\mathbf{C_3}$. However, before routing with RAPID, horizontal surface and subsurface routing is needed to transport runoff from a land surface cell to its corresponding river cell. In the present study, this surface and subsurface routing is done by MODCOU and RAPID replaces only the river modeling of MODCOU.

The connectivity information that already exists between the river cells in the SIM-France river network is used to create the network connectivity matrix $\mathbf{N}$ needed by RAPID and described in Chapter 3..

RAPID uses an automated parameter estimation procedure which, given lateral inflow $\mathbf{Q^e}$ everywhere in the river network, and gage measurements at some locations, determines a best set of parameters based on a square error cost function. As in Chapter

3, the search for optimal vectors of parameters $\mathbf{k}$ and $\mathbf{x}$ is made by determining two multiplying factors λ_k and λ_x such that:

$$\forall j \in [1, 24264] \qquad k_j = \lambda_k \cdot \frac{L_j}{c^0} \quad , \quad x_j = \lambda_x \cdot 0.1 \tag{29}$$

where j is the index of a quad-tree river cell, k_j and x_j are its Muskingum parameters, L_j is the flow distance within a river cell and $c^0 = 1km \cdot h^{-1} = 0.28m \cdot s^{-1}$ is a reference celerity for the flow wave. In this study, the size of the side of each quad-tree river cell was used as an approximation of its flow distance. The value of λ_x is bounded by the interval $[1,5]$ since the Muskingum method is stable only for $x \in [0.1, 0.5]$, as shown in Cunge [1969]. The two scalars λ_k and λ_x are determined such that the corresponding vectors $\mathbf{k}$ and $\mathbf{x}$ minimize the value of an optimization criteria, or cost function. At the end of the optimization procedure, one couple (λ_k, λ_x) is determined for a given part of the network. The values of λ_k and λ_x can be determined for the entire study domain, or for sub-basins. If a sub-basin is located downstream of another sub-basin, observations at a gaging station are used to provide the upstream flow. Therefore, the delineation of sub-basins has to be consistent with the location of available gage measurements.

The optimization procedure uses a line-search algorithm called the Nelder-Mead method [*Nelder and Mead*, 1965] to determine the two scalars λ_k and λ_x.

The use of RAPID within SIM-France allows for flow and volume calculation at each river cell, the river-aquifer interactions are computed more properly, and RAPID allows for the ready inclusion of additional river gages to be used for calibration.

4.4. APPLICATION OF RAPID IN FRANCE

4.4.1. Optimization of RAPID parameters

In order to simplify the optimization procedure and to ensure its repeatability, the parameter estimation of RAPID was run uncoupled from SIM-France. Lateral and groundwater inflow to the river network were obtained from a simulation using the standard version of SIM-France (without RAPID). Daily gage measurements from the French HYDRO database [*SCHAPI*, 2008] were used for the parameter estimation as well as for comparison with daily-averaged flow calculations .

The period of interest of the present study is August 1st 1995 to July 31st 2005. However, the parameter estimation was performed using five months of the first winter (November 1st 1995 to March 31st 1996). As part of the first year (1995-1996) was used for calibration, separate statistical results are presented for 1995-1996 and 1995-2005. RAPID is run using a 15-minute time step and forced with 3-hourly lateral inflow volumes; daily averages of computed discharge are compared with daily observations at gage locations. There are 907 stations within the river network of SIM-France but only 495 of these have daily measurements every day during the first year (August 1st 1995 to July 31st 1996). Amongst the 495 available stations, the best 289 were kept. The criterion used for the selection of the 289 best stations is a Nash efficiency [*Nash and Sutcliffe*, 1970] better than 0.5 in the existing SIM-France model (without RAPID). This selection excludes the gages that are affected either by dams or by water diversions, and thus avoiding unrealistic model parameters due to anthropogenic modification of the river flow. Therefore, the proposed routing scheme is optimized at locations were the previous routing scheme already performed well.

The optimization is first performed on all rivers of the domain, therefore obtaining unique values of λ_k and λ_x for all 24,264 river quad-tree cells. However, such an optimization does not capture the variability between river basins and within sub-basins, due to the various slopes or soil types. Therefore, the optimization procedure was also run independently within the seven main river basins of France shown in Figure 23 and within the twenty sub-basins shown in Figure 24.

In order to limit the effect of the initial state of the system at the beginning of the optimization procedure, the initial flows on 01 November 1995 were estimated using a simple run of RAPID. This estimation was obtained through running the routing model from 01 August to 31 October 1995 with uniform values of λ_k and λ_x over the study domain and initial flows 1 m^3/s for all river cells on 01 August 1995.

The result of a parameter estimation procedure sometimes depends on the initial guess for the parameters. Therefore, three different sets of initial guesses for λ_k and λ_x were used: $(\lambda_k, \lambda_x) = (2,3)$, $(\lambda_k, \lambda_x) = (4,1)$ or $(\lambda_k, \lambda_x) = (1,1)$. The numerical values of these three sets have no particular meaning and serve to start the optimization with a different initial value for $\mathbf{k}$ and $\mathbf{x}$. Each set of initial guesses leads to slightly different results for the optimal λ_k and λ_x. Out of the three sets of optimal λ_k and λ_x that are determined for each sub-basin, only the best is kept. This selection is based on the set of parameters that leads to the smallest value of the optimization cost function.

Once the optimization procedure was completed, SIM-France and RAPID were coupled and run over a 10-year period, from August 1995 to July 2005. In order to compare the overall performance of both routing models on the river network, the Nash efficiency and the root mean square error (RMSE) were calculated for each of the 289 gaging stations. These criteria are sorted and comparisons between the computations of SIM-France and those of SIM-RAPID are shown in Figure 25. The two graphs in Figure

25 do not allow comparing both models at each gaging station since the criteria are sorted, but they depict the overall relative performance of both models. The Nash efficiencies obtained by the original version of SIM-France are higher (an average of 0.04 higher) than those obtained after the addition of RAPID and the RMSEs are very similar.

In its original formulation, the criterion used in the optimization of RAPID is based on a square error cost function ϕ_1. This function is the sum of the square errors between daily measurements $Q_i^g(t)$ and daily-averaged $\overline{Q_i}(t)$ flow computations for several river gaging station i and for everyday of a given period of time $[t_o, t_f]$, as shown in Equation (30).

$$\phi_1(\mathbf{k},\mathbf{x}) = \sum_{t=t_o}^{t=t_f} \sum_{i=1}^{i=289} \left[\frac{\overline{Q_i}(t) - Q_i^g(t)}{f} \right]^2 \tag{30}$$

where the summation is made daily and at river cells with active gaging stations only. t_o and t_f are respectively the first day and last day used for the calculation of ϕ_1. $i \in [1, 289]$ is the index for gaging stations. The model parameter vectors $\mathbf{k}$ and $\mathbf{x}$ are kept constant within the temporal interval $[t_o, t_f]$, and the cost function is calculated several times with different sets of parameters during the optimization procedure. f is a scalar that allows ϕ_1 to be of the order of magnitude of 10^1 which is helpful for automated optimization procedures. It is found that the same fractional error for two stations with different orders of magnitude for river flow influences the cost function differently. A small fractional error on a gaging station with a large flow penalizes the cost function more than the same fractional error on a gaging station with small flow. The Nash

efficiency E is highly influenced by the difference between the model computation and the mean average flow, as shown in Equation (31):

$$E = \frac{\sum_{t=t_o}^{t=t_f} \left[Q_i^g(t) - Q_i(t) \right]^2}{\sum_{t=t_o}^{t=t_f} \left[Q_i^g(t) - \left\langle Q_i^g \right\rangle \right]^2} \tag{31}$$

where $\left\langle Q_i^g \right\rangle$ is the average flow observed at the gaging station i over a long the interval $[t_o, t_f]$. Therefore, the use of ϕ_1 penalizes the Nash efficiency. In order to avoid that the order of magnitude of flow at each gaging station influences their weight in the cost function, a new cost function ϕ_2 is created, as shown in Equation (32).

$$\phi_2(\mathbf{k}, \mathbf{x}) = \sum_{t=t_o}^{t=t_f} \sum_{i=1}^{i=289} \left[\frac{\overline{Q_i}(t) - Q_i^g(t)}{\left\langle Q_i^g \right\rangle} \right]^2 \tag{32}$$

The new cost function ϕ_2 results in the improvements shown in Figure 26 where the Nash efficiencies obtained with RAPID are much higher than with ϕ_1. Overall, the Nash efficiencies and the RMSEs in SIM-RAPID are comparable to those obtained with the routing scheme of the original SIM-France. Therefore, the choice of the cost function is crucial to determining a set of optimal parameters.

In order to estimate the effect of more spatial variability in the parameters of RAPID, the parameter estimation was done on different basins and sub-basins. Figure 27 shows the sorted Nash efficiencies and RMSEs obtained with three degrees of spatial

variability using ϕ_2 as the cost function. These spatial variabilities include "France" which has uniform parameters over the whole domain, "basins" for the 7 river basins of Figure 23 (Adour, Garonne, Loire, Seine, Meuse, Rhône and Hérault) and "sub-basins" where the major river basins have been divided into 20 sub-basins as shown in Figure 24. The increase in spatial variability of parameters improves both the efficiency and the RMSE, but the improvement is limited compared to that triggered by a change in the cost function. The values of parameters k^n and x^n obtained with the parameter estimation procedure using the second cost function are shown in Table 4. The number of gaging stations in a basin can be divided by the number or river cells in the basin to calculate an observability ratio O, as done in Table 4. This ratio ranges from $O = 22$ on the Ardèche River to $O = 1307$ downstream of the Seine River, showing a wide spread in density of observations. The Seine River, of great interest to the French community, has a higher resolution and therefore more river cells in SIM-France than any other basin – all the river cells are of size 1 km – which explains the lower observability ratio.

Figure 28 shows a spatial comparison of results obtained over France. Improvements and degradations of statistical results between SIM-France and SIM-RAPID have no particular spatial patterns. The RMSE and Nash efficiency were also computed for a ten-year simulation and are shown in Figure 29. Overall, the discharge simulated by SIM-France and SIM-RAPID are similar in RMSE and Nash efficiency. This similarity can be explained by the strong dependence of discharge calculations on the lateral inflow forcing which is the same for both river routing schemes. Furthermore, the routing equations used in SIM and SIM-RAPID are comparable (the linear reservoir equation in SIM-FRANCE is a simplified Muskingum equation, given x=0). The addition of RAPID to SIM-France can be regarded as an improvement since RAPID provides with flow and volume of water in all the cells of the river network and provides

flexibility in the number and location of river gages, which was not the case in the original version of SIM-France.

4.4.2. Treatment of dams

RAPID does not have a specific physical model for treatment of dams. However, the model is designed such that observations at gaging stations can easily be substituted for upstream flow. This capability is useful for a gaging station located at the outlet of a dam because the flows discharging from man-made infrastructures reflect human decisions. In France, the quality of flow calculations of the Rhône River (at Beaucaire) is influenced by the dam at the outlet of Lake Geneva. Figure 30 demonstrates the influence of forcing with observations at Pougny (downstream of the dam) on the calculation of flow at the outlet of the Rhône River Basin. The gaging station a Pougny is the outlet of the "Rhône upstream" basin in Figure 24. The first year (August 1st 1995 – July 31st 1996) was used and RAPID was run uncoupled from MODCOU. Forcing with observations at Lake Geneva increases the Nash Efficiency from 0.56 to 0.73 at Beaucaire, the outlet of the Rhône basin.

4.4.3. Improvement of river-aquifer interactions within SIM-RAPID

As discussed previously, the use of isochrone zones within the river network in SIM-France causes inconsistencies in the calculation of the volume of water at the river cell level. Equation (33) shows how river/aquifer interactions are computed in SIM-FRANCE prior to coupling with RAPID.

$$\begin{aligned} &\text{if } H_j - H_0 > 0 \;\Rightarrow\; Q_j^{gw} = TP \cdot \left(H_j - H_0 \right) \\ &\text{if } H_j - H_0 < 0 \;\Rightarrow\; Q_j^{gw} = -\max\left[TP \cdot \left(H_0 - H_j \right), Q^{\min} \right] \end{aligned} \tag{33}$$

where Q_j^{gw} is the volume of water flowing into a given river quad-tree cell j from its corresponding aquifer. H_j is the height of water in the aquifer, and H_0 is the water height in the river reach (considered constant). TP is the streambed conductance, $Q^{\min}$ is a minimal flow allowing aquifer recharge, and was set to a constant value of $Q^{\min} = 0.01 \cdot \mathrm{m}^3/\mathrm{s}$. H_j is computed by MODCOU and H_0 is set based on the digital elevation model (DEM). However, the spatial resolution of DEMs is not usually good enough to provide such information (a depression of the DEM in a neighbor cell could lead to drainage of the aquifer), and thus, some corrections are needed. One of the limitations of the formulation in Equation (33) is that the amount of water that can drain from the river cell to the aquifer is not limited by the quantity of water available in the river cell.

The coupling of SIM with RAPID allows for a finer computation of water volume in river cells, therefore the computation of bi-directional exchanges of water between river and aquifer is improved. The modification applied to river-aquifer interaction is given by Equation (34):

$$
\begin{aligned}
&\text{if } H_j - H_0 > 0 \Rightarrow Q_j^{gw} = TP \cdot \left(H_j - H_0 \right) \\
&\text{if } H_j - H_0 < 0 \Rightarrow \\
&\qquad \text{if } TP \cdot \left(H_0 - H_j \right) > Q_j^{\max} \Rightarrow Q_j^{gw} = -Q_j^{\max} \\
&\qquad \text{if } TP \cdot \left(H_0 - H_j \right) < Q^{\min} \Rightarrow Q_j^{gw} = -Q^{\min} \\
&\qquad \text{if } TP \cdot \left(H_0 - H_j \right) \in \left[Q^{\min}, Q_j^{\max} \right] \Rightarrow Q_j^{gw} = -TP \cdot \left(H_0 - H_j \right)
\end{aligned}
\tag{34}
$$

where $Q_j^{\max}$ is the flow rate corresponding to the drainage of all water available within the river reach. This new formulation differs from Equation (33) because it enables to limit the quantity of water that drains from a river cell to the aquifer by the amount of water available in the given cell.

In the Rhône and Seine River Basins, river-aquifer interactions most often deliver flow from the aquifer to the river network. Hence, the effect of enabling bi-directional exchanges has little effect in the present study. However, in some basins, water quality and water quantity issues related to bi-directional river-aquifer interactions are important. Such is the case, for instance, in the Upper Rhine Aquifer [*Eikenberg, et al.*, 2001].

4.5. CONCLUSIONS

The river routing in SIM-France is done by MODCOU which uses groups of cells called isochrone zones for its computations and does not directly compute flow and volume of water for each cell of its quad-tree river network. The use of isochrones limits the flexibility in the number and location of river gages and generates unstable bi-directional exchanges between rivers and aquifers. The work in this paper presents the replacement of the river routing module in MODCOU by the river network model called RAPID. Information on the network connectivity between the quad-tree river cells of SIM-France is readily available in tables that relate upstream and downstream cells. These tables can be used directly to create the network matrix of RAPID. A ten-year study of river flow in Metropolitan France is presented comparing RAPID and the routing module of MODCOU. An automated procedure for determining optimal model parameters is available in RAPID and various options for the estimation of the parameters are investigated. Sub-basin optimization increases model performance but its influence is much smaller than the choice of the cost function. A cost function was developed that normalizes the square-error between observations at each river gage and RAPID computations by the average flow at the gage. This cost function is found to globally improve the Nash efficiency of computed flow in all gages. We suggest that this is due to the average flow having an influence on the computation of the Nash efficiency. Therefore, the use of an appropriate criterion for quantifying the quality of river flow as the cost function for the optimization procedure helps the betterment of model computations. Overall, the computation obtained with the addition of RAPID are comparable to those of the original river routing module in SIM-France. We consider the addition of RAPID as an improvement since flow and volume of water is directly

computed for each cell of the quad-tree river network. The formulation of RAPID allows for easily substituting observed flows for the upstream calculated flow, which is advantageous when considering a man-made infrastructure as was shown for the Rhône River.

Table 4 Results of optimization procedure using the ϕ_2 cost function

Basin	Sub-basin	Number of river cells	Number of stations	Observability ratio	Optimized λ_k λ_x	Basin	Sub-basin	Number of river cells	Number of stations	Observability ratio	Optimized λ_k λ_x
France	all basin	24264	289	84.0	0.34 0.58	Loire	Loir	670	14	47.9	0.48 0.26
France	all basin	24264	289	84.0	0.34 0.58	Loire	Loire downstream	1763	25	70.5	0.47 0.37
Adour	all basin	666	9	74.0	0.42 2.06	Seine	all basin	5115	39	131.2	0.54 0.04
Garonne	all basin	2985	58	51.5	0.31 0.23	Seine	Seine upstream	2919	29	100.7	0.56 0.72
Garonne	Garonne upstream	558	5	111.6	0.20 0.40	Seine	Oise	889	9	98.8	0.50 2.75
Garonne	Tarn	356	8	44.5	0.16 0.43	Seine	Seine downstream	1307	1	1307.0	0.03 2.48
Garonne	Lot	369	10	36.9	0.34 2.50	Meuse	all basin	832	3	277.3	0.39 0.80
Garonne	Dordogne	431	12	35.9	0.34 0.58	Rhone	all basin	3465	51	67.9	0.30 0.35
Garonne	Garonne downstream	1271	23	55.3	0.41 0.80	Rhone	Saone	1043	32	32.6	0.28 0.36
Loire	all basin	4138	88	47.0	0.43 0.20	Rhone	Ardèche	66	3	22.0	0.23 2.38
Loire	Vienne	706	20	35.3	0.34 0.08	Rhone	Rhone upstream	279	1	279.0	0.50 2.25
Loire	Allier	458	17	26.9	0.30 0.85	Rhone	Rhone downstream	2077	15	138.5	0.38 0.31
Loire	Loire upstream	541	12	45.1	0.46 0.58	Herault	Herault	101	3	33.7	0.31 2.34

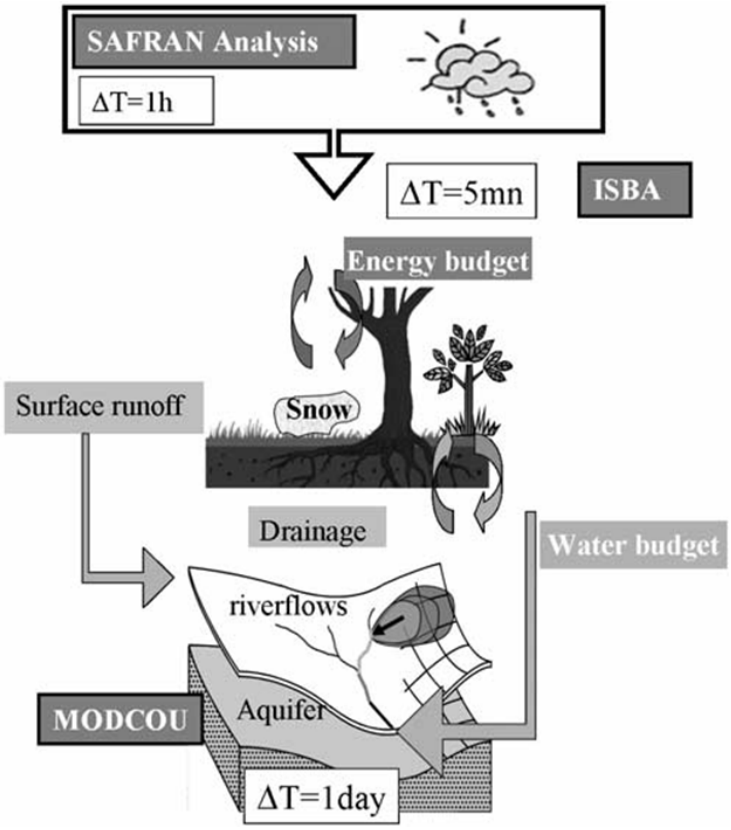

Figure 19 Structure of SIM-France, from Habets et al. [2008]

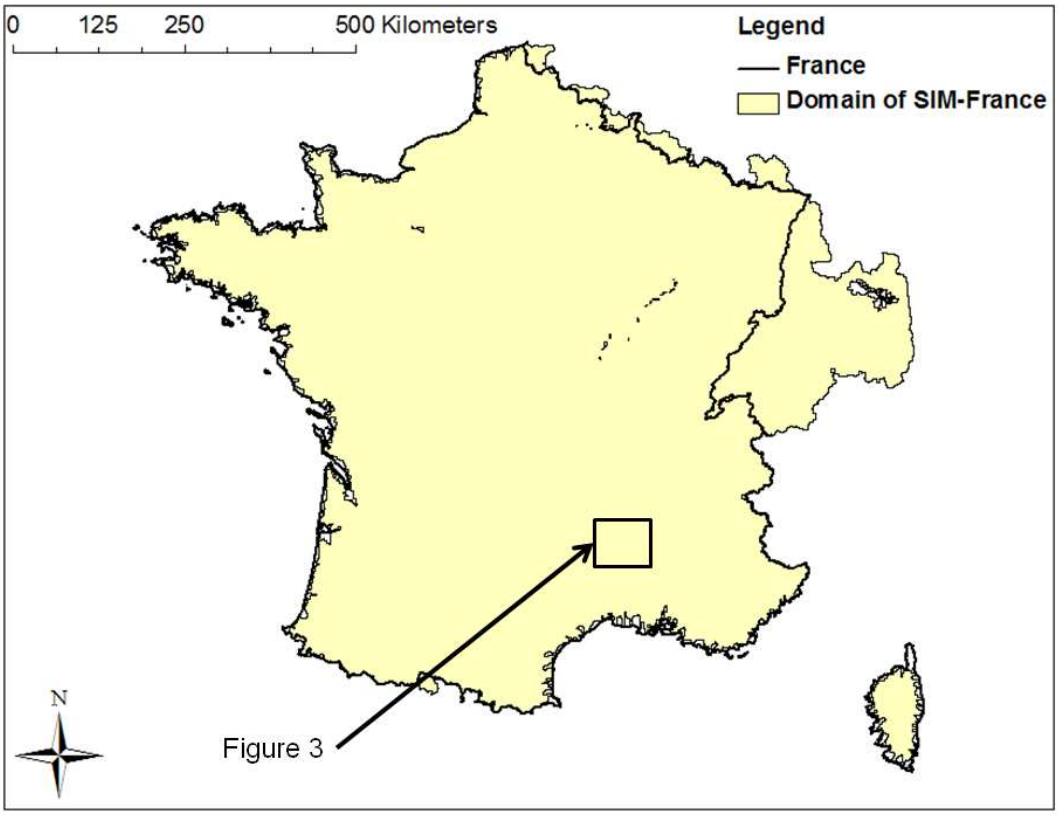

Figure 20 France and computational domain of SIM-France

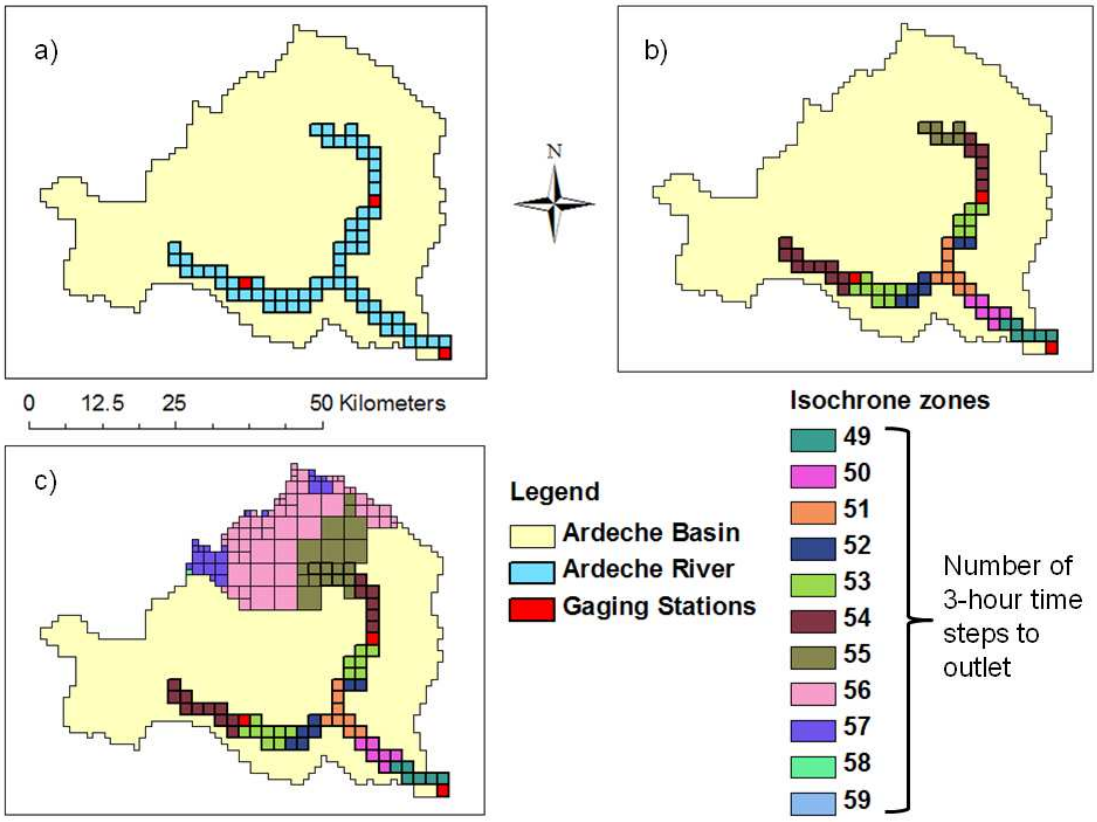

Figure 21 Surface and river isochrone zones in Ardèche Basin

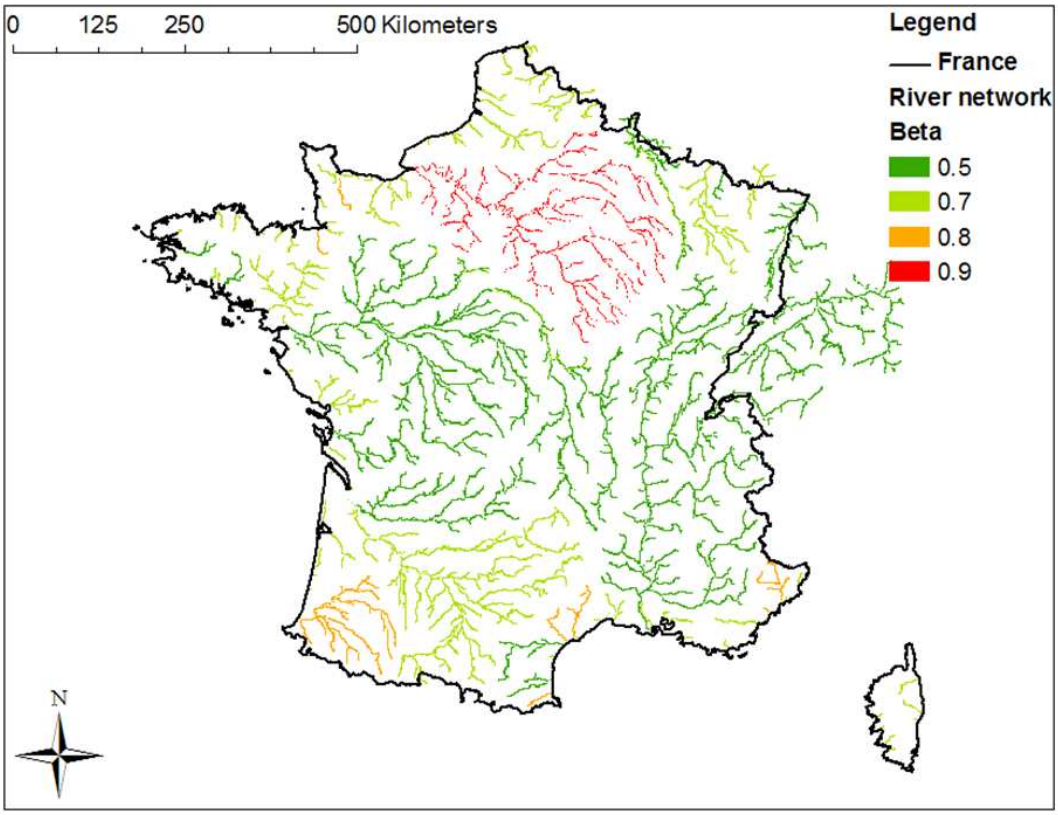

Figure 22 Map of the parameter β used for river routing in SIM-France

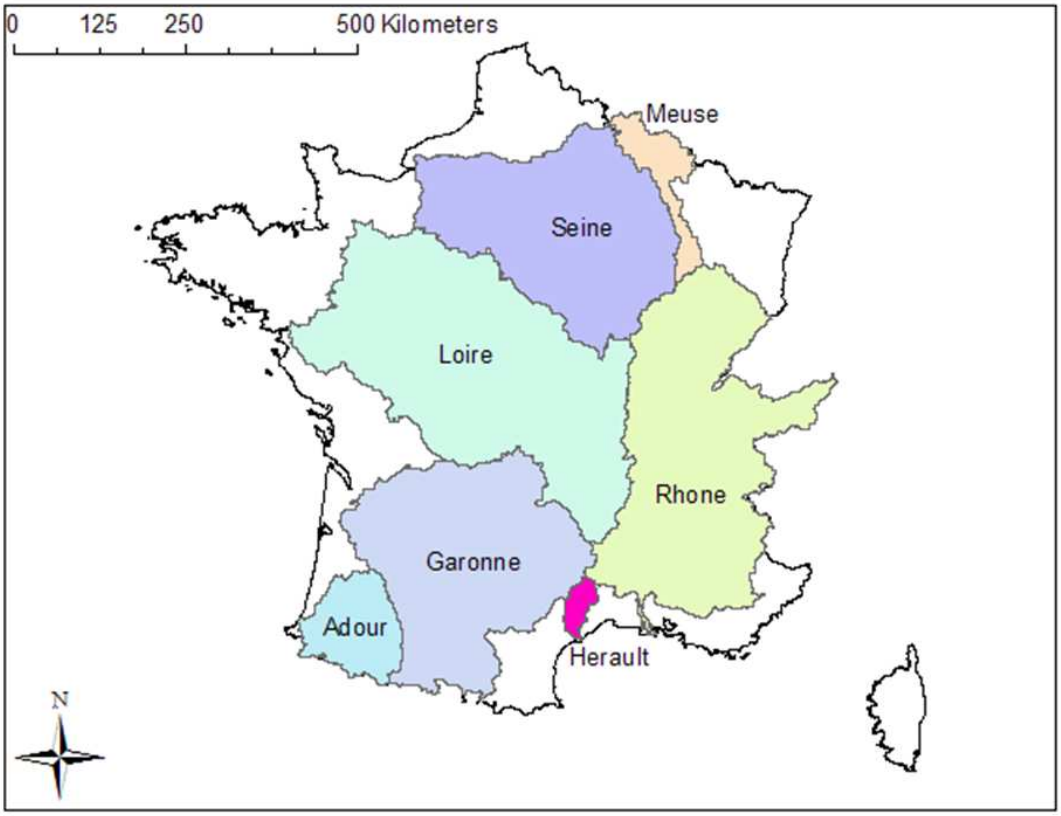

Figure 23 Seven major river basins in SIM-France

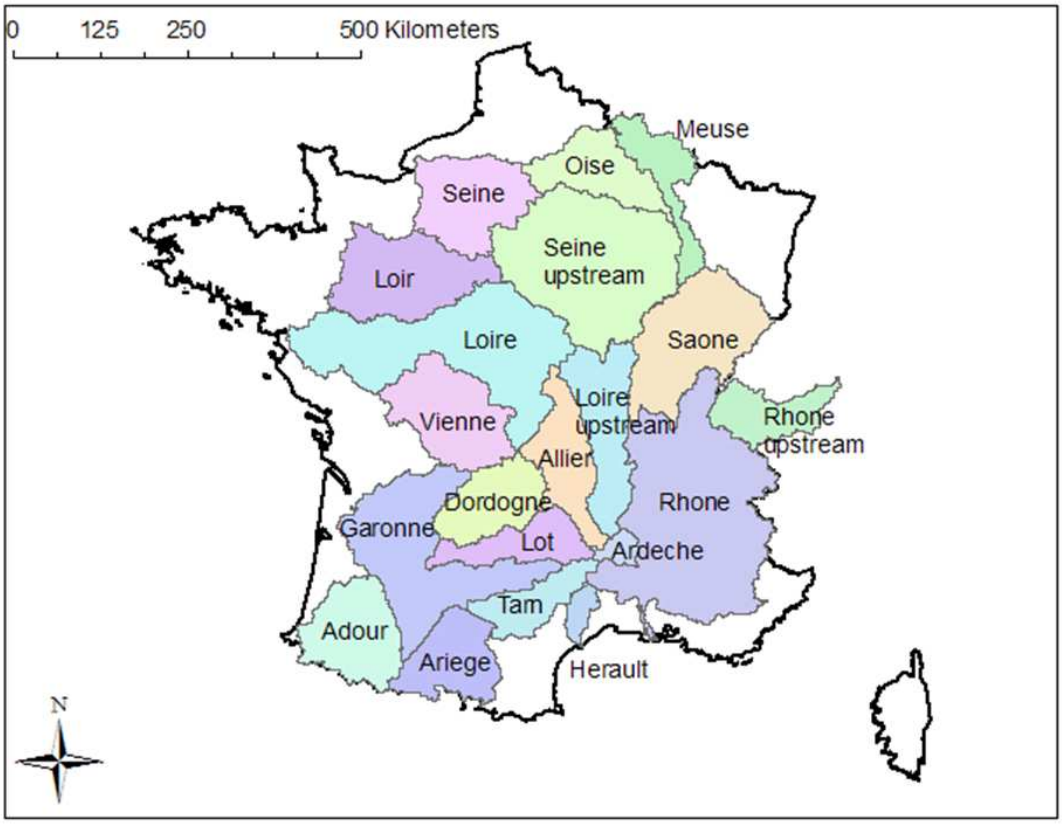

Figure 24 Twenty sub-basins treated independently during optimization of RAPID parameters

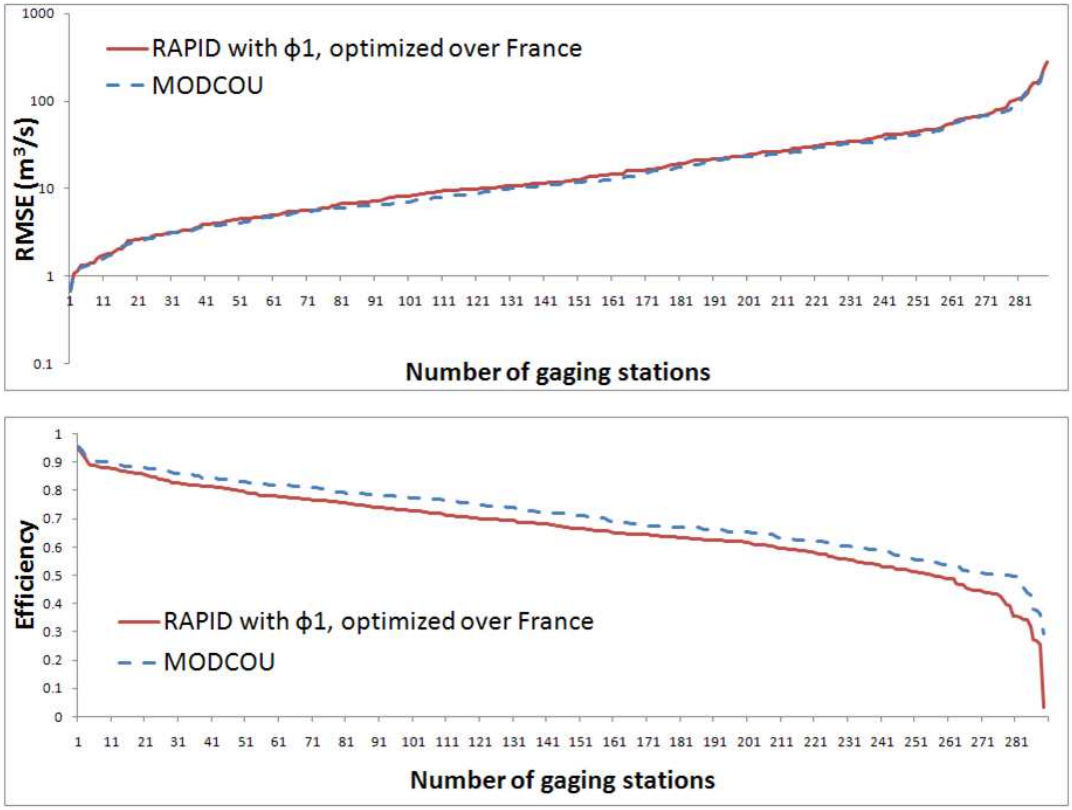

Figure 25 Comparison of sorted RMSE and Nash efficiencies for the year 1995-1996 between SIM-France and RAPID using with parameters obtained with the original cost function ϕ_1

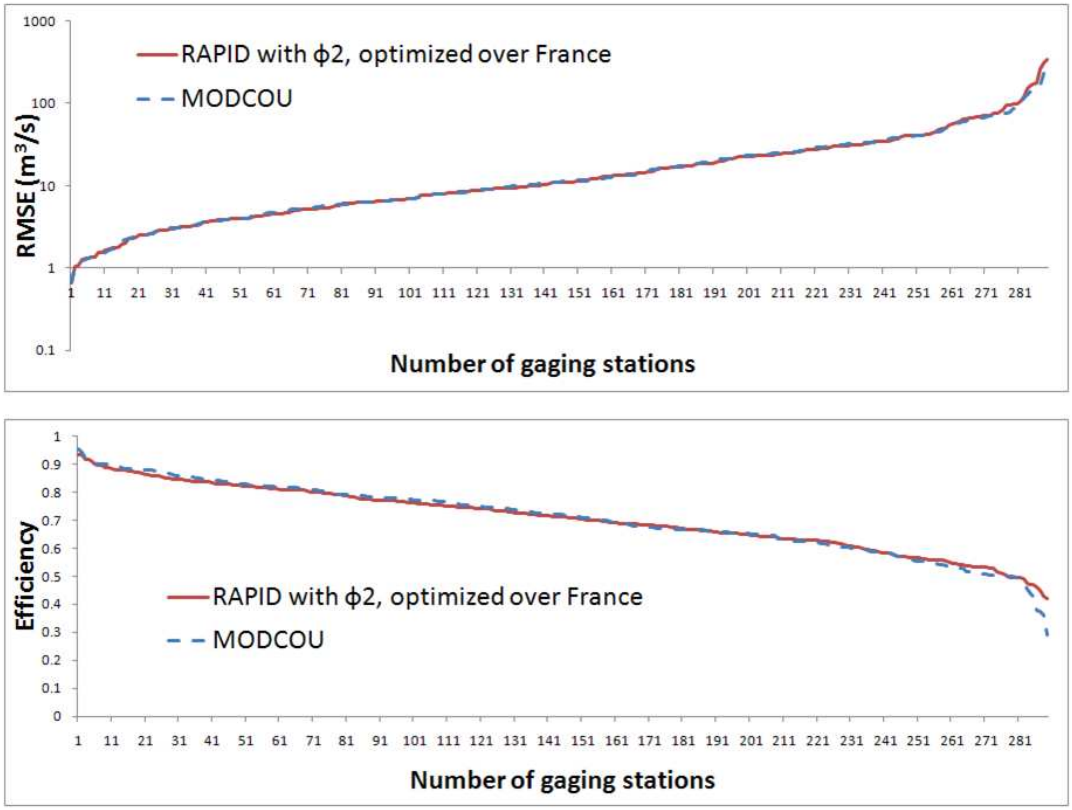

Figure 26 Comparison between RMSE and Nash efficiencies for the year 1995-1996 between SIM-France and RAPID using with parameters obtained with the new cost function ϕ_2

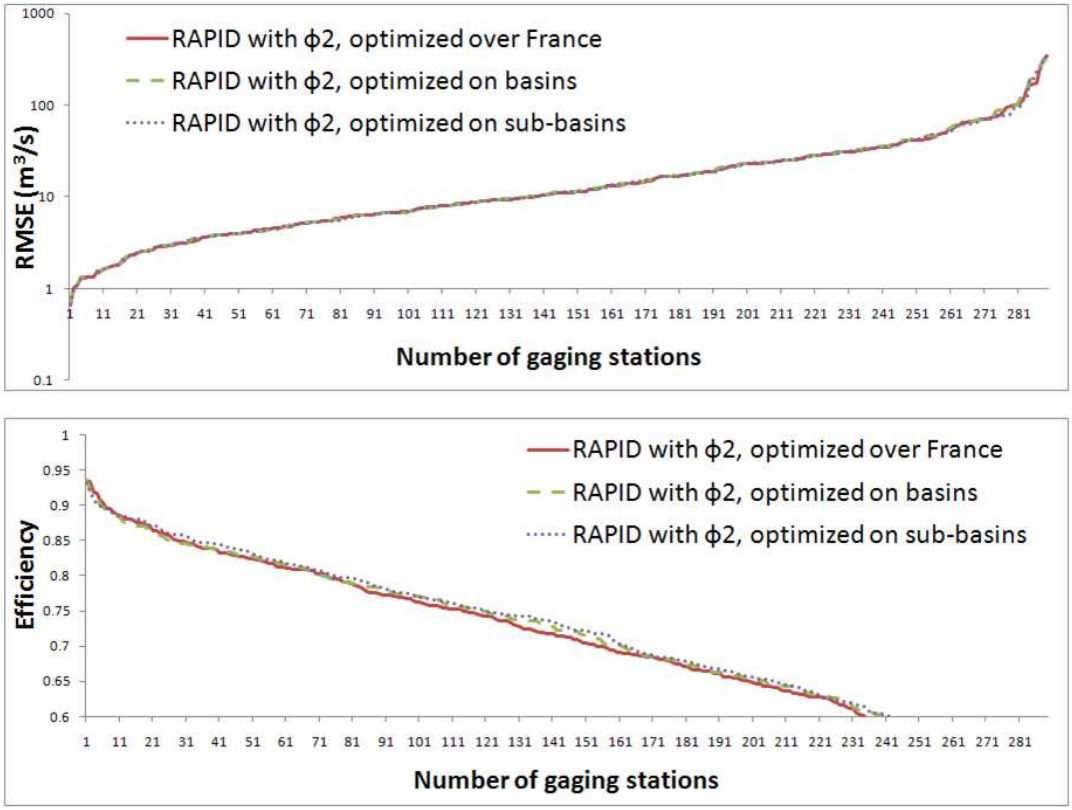

Figure 27 Effect of sub-basin optimization for parameters on RAPID, RMSE and Nash efficiency for the year 1995-1996 using with parameters obtained with the new cost function ϕ_2

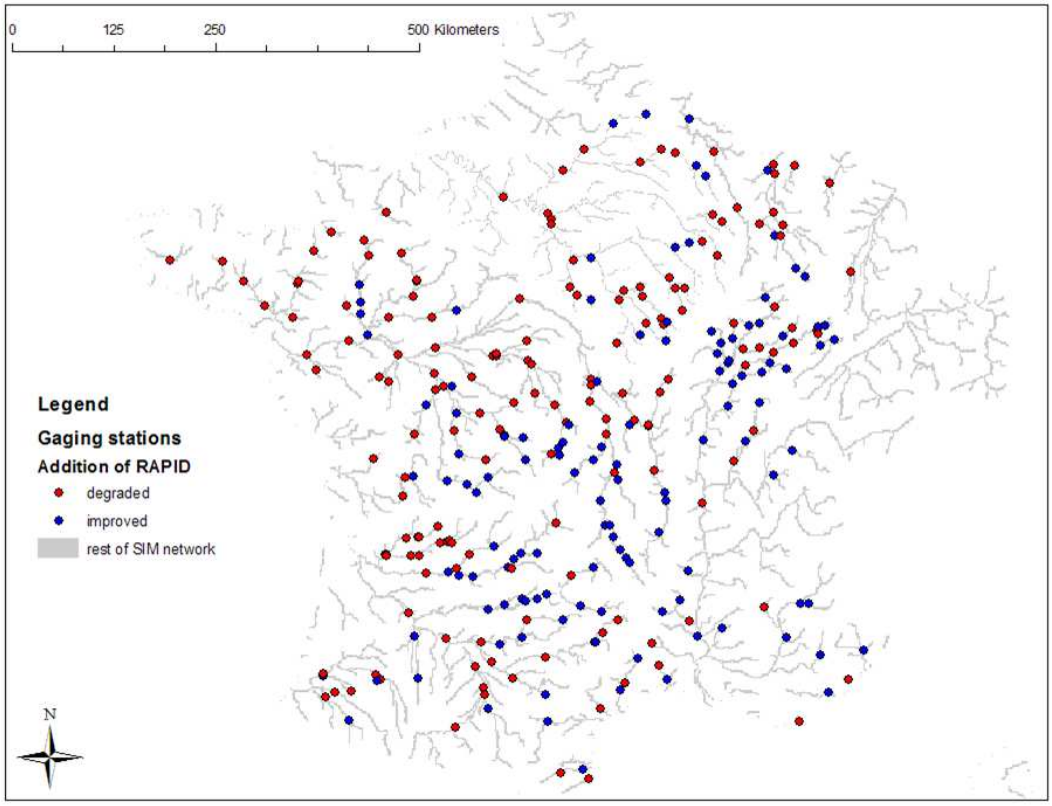

Figure 28 Spatial comparison of results obtained over France with SIM-France and SIM-RAPID for the year 1995-1996 with parameters obtained using the new cost function ϕ_2

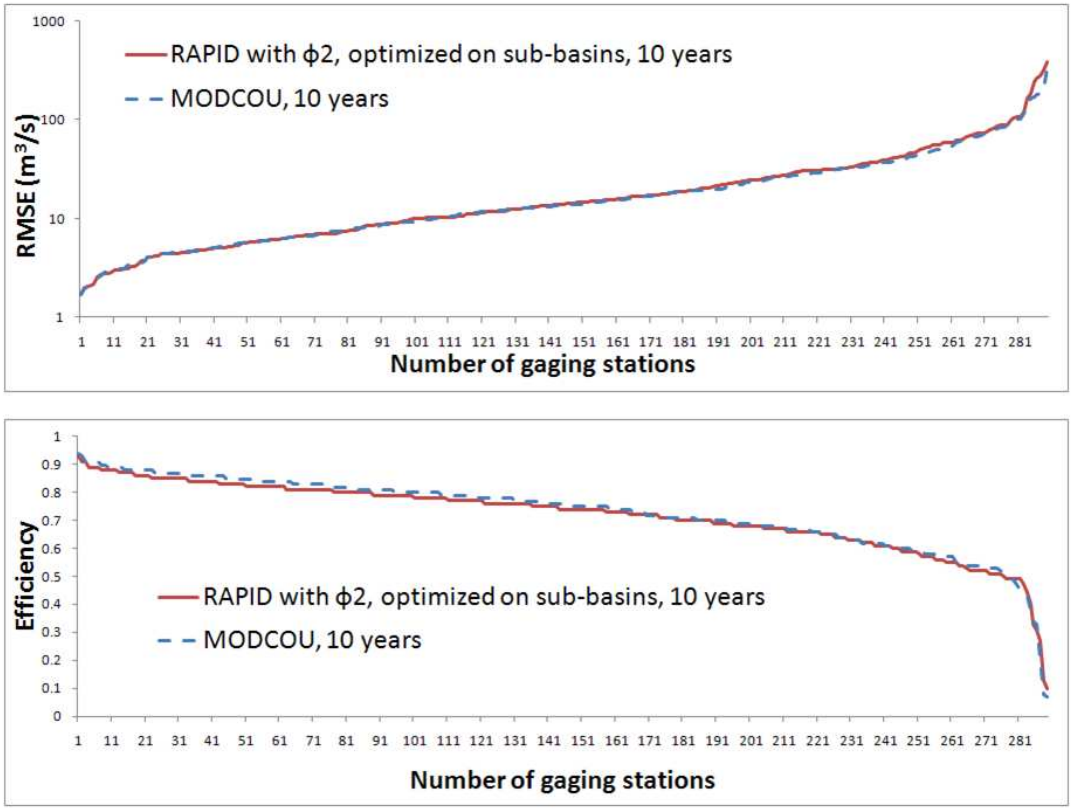

Figure 29 Comparison between RMSE and Nash efficiencies for over ten years (1995-2005) between SIM-France and SIM-RAPID with parameters obtained using the new cost function ϕ_2

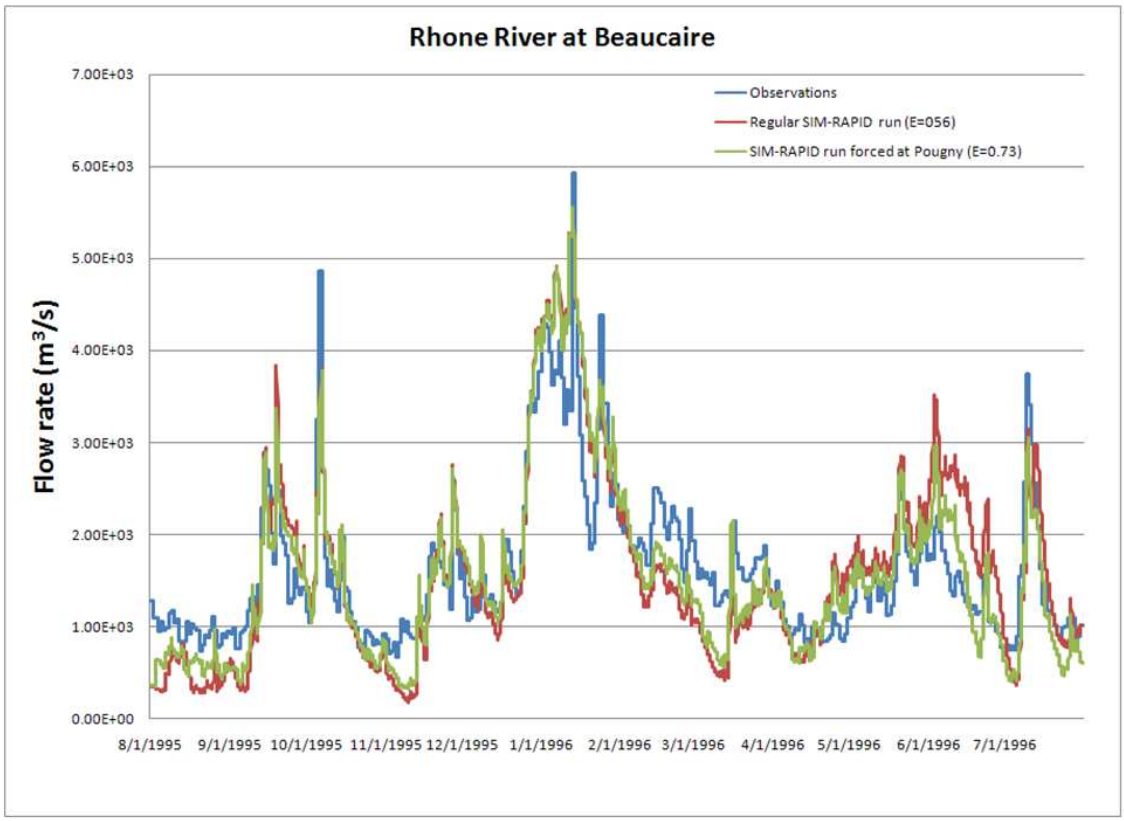

Figure 30 Comparison of SIM-RAPID discharge calculation at the outlet of the Rhône River (at Beaucaire) with and without forcing at the outlet of Lake Geneva (at Pougny).

Chapter 5: Conclusions

5.1. SUMMARY AND CONCLUSIONS

The work presented in this dissertation has addressed the research questions stated in Section 1.2., as summarized in the following.

Chapter 2 presents a catchment "pour point" method as way to use the Noah-distributed model to provide lateral and sub-surface water inflow to an NHDPlus river network. This research identifies different models for the shape of the Earth as a key issue when connecting land surface models with river networks that can lead to errors in distance on the order of 20 km in the North-South direction; proper projections between coordinate systems are needed to avoid these errors. When using the NHDPlus dataset with a land surface model that has surface routing capabilities like Noah-distributed, it is advantageous to use the native coordinate system and fine resolution of NHDPlus and projecting coarser resolution atmospheric data because of the strong connection between DEM, flow direction and flow accumulation grids in NHDPlus.

In Chapter 3, a river network model called RAPID is developed that allows for simultaneous calculation of flow and volume of water in all reaches of a river network. RAPID uses Fortran programming, supercomputers as well as mathematics and optimization libraries that have been especially developed for parallel computing. Although scalability issues are barely tackled, the usefulness of RAPID is demonstrated. In particular, RAPID allows for reconstruction of flow at ungaged river reaches and can be applied over large areas to compute flow for thousands of connected river reaches. The spatial scales used for river routing are an order of magnitude finer than that of previous models; more importantly, actual vector-based river networks extracted from maps are used. RAPID is used with river reaches extracted from an NHDPlus dataset and

framework for computation of river flow on all river reaches of an NHDPlus river network is presented. This framework could be directly applied to larger areas of the 48 United States, or adapted to other domains disposing of comparable vector-based river networks with known connectivity. However, a thorough study of scalability issues is needed before applying such a framework to a continent.

In Chapter 4, the replacement of the river routing of SIM-France by RAPID is presented. River flow is calculated over all the basins of Metropolitan France for a ten-year period and comparable results are obtained with the two routing schemes. However, the flexibility of RAPID and its ability to compute flow and volume of water in all reaches of a network are advantageous. In particular, bi-directional river/aquifer exchanges are made stable. The influence of spatial variability of model parameters as well as the influence of the cost function used for optimization on flow calculations is presented. The latter seems to be more important than the former to improve model results. A criterion for quantifying the global quality of river flow computations using several gages in a river network is presented. Overall, the addition of RAPID to SIM-France can be considered an improvement.

5.2. RECOMMENDATIONS

This research has used stream flow gages that offer both past and near real time observations. However, focus has been on model validation of historical records and not on real time nowcast or forecast of stream flow. In atmospheric science, it is a standard technique that current observations are used to adjust the states of models of numerical weather dynamics [*Romanowicz, et al.*, 2006], and the resulting adjusted models are used to make near-term forecasts. No equivalent capacity to do real-time assimilation of streamflow and groundwater data exists in hydrology even though the capacity to ingest such data is being created, such as through the web services of the CUAHSI Hydrologic

Information System. Therefore, data assimilation is a challenge and future direction that will help improve hydrologic calculations. Although data assimilation of real time measurements is admittedly not one of the goals of the research developed herein, the work presented could serve as the basis for data assimilation of river flow.

References

Abdulla, F. A., D. P. Lettenmaier, E. F. Wood, and J. A. Smith (1996), Application of a macroscale hydrologic model to estimate the water balance of the Arkansas Red River basin, *Journal of Geophysical Research-Atmospheres*, *101*, 7449-7459.

Balay, S., W. D. Gropp, L. C. McInnes, and B. F. Smith (1997), Efficient Management of Parallelism in Object Oriented Numerical Software Libraries, in *Modern Software Tools in Scientific Computing*, edited by E. A. a. A. M. B. a. H. P. Langtangen, pp. 163-202.

Balay, S., K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang (2001), PETSc Web page, Available online at http://www.mcs.anl.gov/petsc.

Balay, S., K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang (2004), PETSc Users Manual, 1-190 pp, Argonne National Laboratory.

Benson, S., L. C. McInnes, J. Moré, T. Munson, and J. Sarich (2007), TAO User Manual (Revision 1.9), Mathematics and Computer Science Division, Argonne National Laboratory, Available online at http://www.mcs.anl.gov/tao.

Berge, C. (1958), Matrice Associée d'un graphe, in *Théorie des Graphes et ses Applications*, edited, pp. 126-128, Dunod, Paris.

Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, *24*, 43-69.

Boone, A., J.-C. Calvet, J. Noilhan, and euml (1999), Inclusion of a Third Soil Layer in a Land Surface Scheme Using the Force-Restore Method, *Journal of Applied Meteorology*, *38*, 1611-1630.

Chen, F., K. Mitchell, J. Schaake, Y. K. Xue, H. L. Pan, V. Koren, Q. Y. Duan, M. Ek, and A. Betts (1996), Modeling of land surface evaporation by four schemes and comparison with FIFE observations, *Journal of Geophysical Research-Atmospheres*, *101*, 7251-7268.

Cunge, J. A. (1969), On the subject of a flood propagation computation method (Muskingum method), *Journal of Hydraulic Research*, *7*, 205-230.

David, C. H., D. J. Gochis, D. R. Maidment, W. Yu, D. N. Yates, and Z.-L. Yang (2009), Using NHDPlus as the Land Base for the Noah-distributed model, *Transactions in GIS*, *(accepted for publication)*.

De Roo, A., B. Gouweleeuw, and J. Thielen (2003), Development of a European flood forecasting system, *International Journal of River Basin Management*, *1*, 49-59.

De Roo, A. P. J., C. G. Wesseling, and W. P. A. Van Deursen (2000), Physically based river basin modelling within a GIS: the LISFLOOD model, *Hydrological Processes*, *14*, 1981-1992.

Dickinson, R. E., K. W. Oleson, G. Bonan, F. Hoffman, P. Thornton, M. Vertenstein, Z. L. Yang, and X. B. Zeng (2006), The Community Land Model and its climate statistics as a component of the Community Climate System Model, *Journal of Climate*, *19*, 2302-2324.

Dirmeyer, P. A., A. J. Dolman, and N. Sato (1999), The pilot phase of the Global Soil Wetness Project, *Bulletin of the American Meteorological Society, 80*, 851-878.

Dongarra, J., D. Walker, E. Lusk, B. Knighten, M. Snir, A. Geist, S. Otto, R. Hempel, E. Lusk, W. Gropp, J. Cownie, T. Skjellum, L. Clarke, R. Littlefield, M. Sears, S. Hussleder man, E. Anderson, S. Berryman, J. Feeney, D. Frye, L. Hart, A. Ho, J. Kohl, P. Madams, C. Mosher, P. Pierce, E. Schikuta, R. G. Voigt, R. Babb, R. Bjornson, V. Fernando, I. Glendinning, T. Haupt, C. T. H. Ho, S. Krauss, A. Mainwaring, D. Nessett, S. Ranka, A. Singh, D. Weeks, J. Baron, N. Doss, S. Fineberg, A. Greenberg, D. Heller, G. Howell, B. Leary, O. McBryan, P. Pacheco, P. Rigsbee, A. Sussman, S. Wheat, E. Barszcz, A. Elster, J. Flower, R. Harrison, T. Henderson, J. Kapenga, A. Maccabe, P. McKinley, H. Palmer, A. Robison, R. Tomlinson, and S. Zenith (1994), Special Issue - MPI - a Message-Passing Interface Standard, *International Journal of Supercomputer Applications and High Performance Computing, 8*, 159-416.

Durand, Y., E. Brun, L. Mérindol, G. Guyomarc'h, B. Lesaffre, and E. Martin (1993), A meteorogical estimation of relevant parameters for snow models, *Annals of Glaciology, 18*, 65-71.

Eikenberg, J., A. Tricca, G. Vezzu, P. Stille, S. Bajo, and M. Ruethi (2001), Ra-228/Ra-226/Ra-224 and Sr-87/Sr-86 isotope relationships for determining interactions between ground and river water in the upper Rhine valley, *Journal of Environmental Radioactivity, 54*, 133-162.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the National

Centers for Environmental Prediction operational mesoscale Eta model, *Journal of Geophysical Research-Atmospheres*, *108*, 1-16.

Fread, D. L. (1993), Flow Routing, in *Handbook of Hydrology*, edited by D. R. Maidment, pp. 10.17-10.18, McGraw-Hill, New York.

Gates, W. L. (2004), Derivation of the equations of atmospheric motion in oblate spheroidal coordinates, *Journal of the Atmospheric Sciences*, *61*, 2478-2487.

Gochis, D. J., and F. Chen (2003), Hydrological Enhancements to the Community Noah Land Surface Model, available online at http://www.ucar.edu/library/collections/technotes/technotes.jsp.

Guan, H., J. L. Wilson, and H. Xie (2009), A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain, *Journal of Hydrology*, (accepted).

Gunduz, O., and M. M. Aral (2005), River networks and groundwater flow: a simultaneous solution of a coupled system, *Journal of Hydrology*, *301*, 216-234.

Habets, F., J. Noilhan, C. Golaz, J. P. Goutorbe, P. Lacarrere, E. Martin, C. Ottle, and D. Vidal-Madjar (1999a), The ISBA surface scheme in a macroscale hydrological model applied to the Hapex-Mobilhy area - Part I: Model and database, *Journal of Hydrology*, *217*, 75-96.

Habets, F., J. Noilhan, C. Golaz, J. P. Goutorbe, P. Lacarrere, E. Leblois, E. Ledoux, E. Martin, C. Ottle, and D. Vidal-Madjar (1999b), The ISBA surface scheme in a

macroscale hydrological model applied to the Hapex-Mobilhy area - Part II: Simulation of streamflows and annual water budget, *Journal of Hydrology*, *217*, 97-118.

Habets, F., P. Etchevers, C. Golaz, E. Leblois, E. Ledoux, E. Martin, J. Noilhan, and C. Ottle (1999c), Simulation of the water budget and the river flows of the Rhone basin, *Journal of Geophysical Research-Atmospheres*, *104*, 31145-31172.

Habets, F., A. Boone, J. L. Champeaux, P. Etchevers, L. Franchisteguy, E. Leblois, E. Ledoux, P. Le Moigne, E. Martin, S. Morel, J. Noilhan, P. Q. Segui, F. Rousset-Regimbeau, and P. Viennot (2008), The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *Journal of Geophysical Research-Atmospheres*, *113*.

Hellweger, F., and D. Maidment (1997), AGREE - DEM Surface Reconditioning System, available online at http://www.ce.utexas.edu/prof/maidment/gishydro/ferdi/research/agree/agree.html.

Henderson-Sellers, A., Z. L. Yang, and R. E. Dickinson (1993), The Project for Intercomparison of Land-Surface Parameterization Schemes, *Bulletin of the American Meteorological Society*, *74*, 1335-1349.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004), Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resources Research*, *40*, 1-23.

Julien, P. Y., B. Saghafian, and F. L. Ogden (1995), Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff, *Water Resources Bulletin*, *31*, 523-536.

Leblanc, D., and J. P. Villeneuve (1978), Algorithme de schématisation des écoulements d'un bassin versant 1-55 pp, Institut national de la recherche scientifique, Québec.

Leblois, E., and E. Sauquet (2000), Grid elevation models in hydrology. Part 2: HydroDem technical note, 80 pp, Cemagref, Lyon.

Ledoux, E., G. Girard, G. de Marsily, J. P. Villeneuve, and J. Deschenes (1989), Spatially Distributed Modeling: Conceptual Approach, Coupling Surface Water and Groundwater, in *Unsaturated Flow in Hydrologic Modeling Theory and Practice*, edited by H. J. Morel-Seytoux, pp. 435-454, Kluwer Academic Publishers.

Lehner, B., K. Verdin, and A. Jarvis (2006), HydroSHEDS Technical Documentation, available online at http://hydrosheds.cr.usgs.gov.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A Simple Hydrologically Based Model of Land-Surface Water and Energy Fluxes for General-Circulation Models, *Journal of Geophysical Research-Atmospheres*, *99*, 14415-14428.

Liang, X., D. P. Lettenmaier, and E. F. Wood (1996), One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, *Journal of Geophysical Research-Atmospheres*, *101*, 21403-21422.

Lohmann, D., R. NolteHolube, and E. Raschke (1996), A large-scale horizontal routing model to be coupled to land surface parametrization schemes, *Tellus Series a-Dynamic Meteorology and Oceanography*, *48*, 708-721.

Lohmann, D., D. P. Lettenmaier, X. Liang, E. F. Wood, A. Boone, S. Chang, F. Chen, Y. J. Dai, C. Desborough, R. E. Dickinson, Q. Y. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E. Mitchell, O. N. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. P. Shao, A. B. Shmakin, D. Verseghy, K. Warrach, P. Wetzel, Y. K. Xue, Z. L. Yang, and Q. C. Zeng (1998a), The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes, *Global and Planetary Change*, *19*, 161-179.

Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998b), Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *43*, 131-141.

Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998c), Regional scale hydrology: II. Application of the VIC-2L model to the Weser River, Germany, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *43*, 143-158.

Lohmann, D., K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A. Cosgrove, J. Sheffield, Q. Y. Duan, L. F. Luo, R. W. Higgins, R. T. Pinker, and J. D. Tarpley (2004), Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project, *Journal of Geophysical Research-Atmospheres*, *109*, 1-22.

Lyon, S. W., F. Dominguez, D. J. Gochis, N. A. Brunsell, C. L. Castro, F. K. Chow, Y. Fan, D. Fuka, Y. Hong, P. A. Kucera, S. W. Nesbitt, N. Salzmann, J. Schmidli, P. K. Snyder, A. J. Teuling, T. E. Twine, S. Levis, J. D. Lundquist, G. D. Salvucci, A. M.

Sealy, and M. T. Walter (2008), Coupling terrestrial and atmospheric water dynamics to improve prediction in a changing environment, *Bulletin of the American Meteorological Society*, *89*, 1275-1279.

Maidment, D. R., F. Olivera, A. Calver, A. Eatherall, and W. Fraczek (1996), Unit hydrograph derived from a spatially distributed velocity field, *Hydrological Processes*, *10*, 831-844.

Maurer, E. P., G. M. O'Donnell, D. P. Lettenmaier, and J. O. Roads (2001), Evaluation of the land surface water budget in NCEP/NCAR and NCEP/DOE reanalyses using an off-line hydrologic model, *Journal of Geophysical Research-Atmospheres*, *106*, 17841-17862.

McCarthy, G. T. (1938), The Unit Hydrograph and Flood Routing, paper presented at Conference of the North Atlantic Division, U.S. Engineer Department, New London, Connecticut, on June 24, 1938.

Mitchell, K. (2005), The Community Noah Land Surface Model User's Guide, available online at ftp://ftp.emc.ncep.noaa.gov/mmb/gcp/ldas/noahlsm/ver_2.7.1/.

Moore, R. B., C. M. Johnston, K. W. Robinson, and J. R. Deacon (2004), Estimation of Total Nitrogen and Phosphorus in New England Streams Using Spatially Referenced Regression Models, USDI, USGS, available online at http://pubs.usgs.gov/sir/2004/5012/.

Moritz, H. (1980), Geodetic Reference System 1980, *Journal of Geodesy*, *54*, 395-405.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I -- A discussion of principles, *Journal of Hydrology*, *10*, 282-290.

Nelder, J. A., and R. Mead (1965), A Simplex Method for Function Minimization, *The Computer Journal*, *7*, 308-313.

Ngo-Duc, T., T. Oki, and S. Kanae (2007), A variable streamflow velocity method for global river routing model: model description and preliminary results, *Hydrol. Earth Syst. Sci. Discuss.*, *4*, 4389-4414.

Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow simulation for continental-scale river basins, *Water Resources Research*, *33*, 711-724.

Niu, G. Y., Z. L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *Journal of Geophysical Research-Atmospheres*, *112*, 1-14.

Niu, G. Y., Z. L. Yang, K. E. Mitchell, F. Chen, M. Ek, M. Barlage, L. Longuevergne, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y.-L. Xia (2009), The Community Noah Land Surface Model with Multi-Physics Options, *Journal of Geophysical Research-Atmospheres*, (submitted).

NOAA (2005), NWSRFS Users Manual, available online at http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/htm/formats.htm.

Noilhan, J., and S. Planton (1989), A Simple Parameterization of Land Surface Processes for Meteorological Models, *Monthly Weather Review*, *117*, 536-549.

Oki, T., and Y. C. Sud (1998), Design of Total Runoff Integrating Pathways (TRIP) - A Global River Channel Network, *Earth Interactions*, *2*, 1-35.

Oki, T., Y. Agata, S. Kanae, T. Saruhashi, D. W. Yang, and K. Musiake (2001), Global assessment of current water resources using total runoff integrating pathways, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *46*, 983-995.

Olivera, F., and D. Maidment (1999), Geographic information systems (GIS)-based spatially distributed model for runoff routing, *Water Resources Research*, *35*, 1155-1164.

Olivera, F., J. Famiglietti, and K. Asante (2000), Global-scale flow routing using a source-to-sink algorithm, *Water Resources Research*, *36*, 2197-2207.

Quintana-Segui, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, and S. Morel (2008), Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *Journal of Applied Meteorology and Climatology*, *47*, 92-107.

Quintana SeguÃ-, P., E. Martin, F. Habets, and J. Noilhan (2009), Improvement, calibration and validation of a distributed hydrological model over France, *Hydrol. Earth Syst. Sci.*, *13*, 163-181.

Romanowicz, R. J., P. C. Young, and K. J. Beven (2006), Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom, *Water Resources Research*, *42*, 1-12.

Sahoo, G. B., C. Ray, and E. H. De Carlo (2006), Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream, *Journal of Hydrology*, *327*, 94-109.

SCHAPI (2008), Banque HYDRO, Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations available online at http://www.hydro.eaufrance.fr/index.php.

Schwarz, C., and E. Wade (1990), The North American datum of 1983: Project methodology and execution, *Journal of Geodesy*, *64*, 28-62.

Shuttleworth, W. J. (1988), Macrohydrology -- The new challenge for process hydrology, *Journal of Hydrology*, *100*, 31-56.

Smith, M. B., D. J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004), The distributed model intercomparison project (DMIP): motivation and experiment design, *Journal of Hydrology*, *298*, 4-26.

Todini, E. (2007), A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach (vol 11, pg 1645, 2007), *Hydrology and Earth System Sciences*, *11*, 1783-1783.

USEPA, and USGS (2007), NHDPlus User Guide, available online at http://www.horizon-systems.com/nhdplus/documentation.php.

Van Sickle, J. (2004), *Basic GIS Coordinates*, CRC Press LLC, Boca Raton, FL.

Welles, E., S. Sorooshian, G. Carter, and B. Olsen (2007), Hydrologic verification - A call for action and collaboration, *Bulletin of the American Meteorological Society*, *88*, 503-511.

Wetzel, S. (1994), A hydrological model for predicting the effects of climate change, 85 pp, Princeton University, Princeton.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier (1994), A Distributed Hydrology-Vegetation Model for Complex Terrain, *Water Resources Research*, *30*, 1665-1679.

Yang, Z. L. (2004), Modeling land surface processes in short-term weather and climate studies, in *Observations, Theory, and Modeling of Atmospheric Variability*, edited by X. Zhu, pp. 288-313, World Scientific Publishing Corporation, Singapore.

Vita

Cédric Hervé David was born in Marseille, France on December 1, 1981, the son of Brigitte Marie Jarnot and Hervé Jean David. Cédric received a Master of Science in General Engineering from Ecole Centrale de Lille (Lille, France) in August 2004, and moved to the United States of America. In September 2004, he entered the Graduate School the University of Texas at Austin and received a Master of Science in Environmental and Water Resources Engineering in May 2006.

Permanent address: 6802 Kings PT W, Austin, TX 78723, USA

This dissertation was typed by the author.